BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12Q 1/68 — A2

(11) International Publication Number: WO 00/56923

(43) International Publication Date: 28 September 2000 (28.09.00)

(21) International Application Number: PCT/GB00/01128

(22) International Filing Date: 24 March 2000 (24.03.00)

(30) Priority Data:

9906833.0	24 March 1999 (24.03.99)	GB
9927520.8	23 November 1999 (23.11.99)	GB

(71) Applicant *(for all designated States except US)*: CLATTERBRIDGE CANCER RESEARCH TRUST [GB/GB]; J. K. Douglas Laboratories, Clatterbridge Hospital, Bebington, Wirral, Cheshire CH63 4JY (GB).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: SIBSON, Ross [GB/GB]; One Castlehill Farm Barn, Castlehill, Kingswood, Frodsham, Cheshire WA6 6JS (GB).

(74) Agent: MCNEIGHT & LAWRENCE; Regent House, Heaton Lane, Stockport, Cheshire SK4 1BS (GB).

(81) Designated States: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published

Without international search report and to be republished upon receipt of that report.

(54) Title: GENETIC ANALYSIS

(57) Abstract

The present invention relates to genetic analysis of nucleic acids, particularly the analysis of the structure and/or sequence of polynucleotides. The invention also relates to the field of oligonucleotide probes, particularly probes in the form of libraries of oligonucleotide fragments. The invention further concerns the construction of oligonucleotide libraries and the methods of their use in the elucidation of structural or sequence information of sample sequences.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

Genetic Analysis

The present invention relates to genetic analysis of nucleic acids, particularly the analysis of the structure and/or sequence of polynucleotides. The invention also relates to the field of oligonucleotide probes, particularly probes in the form of libraries of oligonucleotide fragments. The invention further concerns the construction of oligonucleotide libraries and the methods of their use in the elucidation of structural or sequence information of sample sequences.

In molecular biology there are circumstances when there is a need to determine features associated with the sequence of a nucleic acid, whether deoxyribonucleic acid or ribonucleic acid, single stranded, double stranded, part single stranded or part double stranded. Single stranded regions of polynucleotides may fold so that Watson and Crick base pairing occurs between parts of the single stranded regions thereby resulting in tertiary structure. These tertiary structures may have biological significance and research efforts focus on an identification and characterisation of the tertiary structure via an analysis of the nucleic acid sequence involved. Certain bases in a polynucleotide may be chemically modified and again if there is biological significance attaching to such modified bases then researchers will seek to identify and characterise modifications in polynucleotides.

Inherent features of a polynucleotide may be its sequence *per se*, ie the order in which bases occur when moving from one part of the sequence to another (or part of the same); some consequence of the sequence, for example amino acid coding capacity; an alteration in the sequence (for example a mutation); a site

of cleavage or the point of an interaction with some other chemical which may include proteins or other nucleic acids.

Perhaps one of the more straightforward analyses which is required from time to time is to determine the position and identity a single base difference between two otherwise identical nucleotide sequences, for example a single nucleotide polymorphism (Wang D G *et al* (1998) Science 280:1077-1082). Single nucleotide polymorphisms are single base differences found at a particular point in the same otherwise identical sequences from different individuals of an interbreeding population or from the same point in a particular chromosome pair of an individual. Single nucleotide polymorphisms are important in biology because they can often be linked to heritable traits the most important of which are the inherited components of common diseases, for example arteriosclerosis, cancer and diabetes.

By way of background, the two strands of a double stranded nucleic acid are held together by hydrogen bonding between the purine and pyrimidine bases that are correctly paired according to Watson and Crick base pairing rules. Strands which are correctly base paired together are said to be complementary. Hydrogen bonds are weaker than other chemical bonds so that heat or other denaturing agents which disrupt base pairing cause double stranded nucleic acids to separate into their component single strands. Denaturation is reversible and double strands reform when complementary single strands reanneal together in a process called nucleic acid hybridisation. Sequences do not have to be perfectly complementary in order to hybridise together. So long as there are a sufficient number of correctly paired bases linked by hydrogen bonds then the two strands are prevented from moving apart. Points at which double stranded sequences are not correctly paired are said to be mismatched. Regions

of mismatch can range from a single base to extensive regions depending upon the prevailing conditions of denaturation. Under certain conditions strands may still hybridise despite mismatching.

Assays dependent on detecting an oligonucleotide mismatch have conventionally been used for detecting single base alterations at known sites in a sequence of interest. Typically, mismatched oligonucleotides are not able to hybridise to a target sequence under conditions of high stringency.

Oligonucleotides which hybridise perfectly to mutant or wild type sequences respective have been used to determine whether the mutant or wild type sequence is present in a sample. The respective oligonucleotides are differentially labelled. (See for example, Holland P M *et al* (1991) PNAS 88(16):7276-7280; Livak K J *et al.* (1995) PCR Methods Appl 4(6):357-362; Mergny J L *et al* (1994) Nucl Acids Res 22(6):920-928; Livak K J *et al.* (1995) Nat Genet 9(4):341-342; Heinonen P *et al.* (1997) 43(7):1142-1150).

The Taqman methods are typically used during the course of or after amplification of the sample sequence, for example by the polymerase chain reaction or PCR which is well known to those skilled in the art.

There are a number of existing methods which are able to amplify preferentially a given sequence, thus serving to identify the presence of that sequence amongst a number of other similar sequences in a sample. The sequence to be amplified preferentially may be a mutant sequence and the sample may contain mutant and wild type sequences.

The operation and success of these methods relies on the fact that DNA polymerases have fidelity. The DNA polymerases are able to detect the presence of mismatched bases in annealed strands - such mismatches inhibit the DNA polymerases from catalysing further polymerisations of nucleotides at those positions. The basis of the Amplification Refractory Mutation System (ARMS) is that oligonucleotides with a mismatched 3' residue will not function as primers in the PCR under certain conditions (see Newton C R *et al.* (1989) Nucleic Acids Res 17(7):2503-2516). By siting mismatching bases at appropriate positions in PCR primers, such primers are thus able to discriminate between alternative versions of template strands. Only those primer-template hybrids which do not involve a mismatch are able to be amplified and thereby be identifiable.

Similarly, the oligonucleotide ligation amplification (OLA) assay is a method of amplifying DNA that uses oligonucleotides (Baron H *et al.* (1996) Nat Biotechnol 14(10):1279-1282). Ligation is the covalent linking of the 3' end of one nucleic acid to the 5' end of another nucleic acid, catalysed by the enzyme ligase. Ligase requires a free 3' hydroxyl on a 3' end and a 5' phosphate on a 5' end. Two oligonucleotides can be prepared so that when they are annealed to a template the 5' end of one is directly adjacent to the 3' end of the other; in other words the two ends are juxtaposed. Template dependent ligation between the juxtaposed oligonucleotides can then take place. At its simplest the OLA assay requires 4 oligonucleotides which together correspond to each strand of the sequence to be amplified. They are prepared two per complementary strand such that their sites of hybridisation brings ends of the oligonucleotides together into juxtaposition. The ligation products from one strand can then serve as the hybridisation sites for the oligonucleotides from the complementary strand. Denaturation of the resultant ligation products following ligation allows the reaction to be repeated in a cyclical fashion with resultant exponential increase

in the ligated products. Template sample sequences can be distinguished from one another if they possess a base difference in the regions of hybridisation with the oligonucleotides because this will reduce the effectiveness of hybridisation and/or prevent ligation from occurring. Ligase has a fidelity which means it can recognise when nucleic acids are not correctly paired and this reduces the rate at which it catalyses the covalent joining of the ends of juxtaposed oligonucleotides. Inhibition of ligation is greatest if the mismatched bases are found at the point where the oligonucleotides are juxtaposed.

Use of oligonucleotide based methods including ARMS and the OLA assay to detect multiple mutations in a sequence is hampered generally by the limited number of ways in which oligonucleotides can be differentially labelled for detection. This restricts the applicability of the aforementioned methods to the detection of likely and known differences between sequences.

In the case of methods used for identifying and screening differences between polynucleotides which remain unknown, some general methods do exist. In general, these are indirect and rely upon the presence of a sequence variant to bring about differences in mobility between polynucleotides during gel electrophoresis of the same. Temperature gradient gel electrophoresis (TGGE) uses a temperature gradient in parallel with a voltage gradient and this in the denaturation of denature sequences during their electrophoresis (Reisner D *et al.* (1992) Electrophoresis 13(9-10): 632-636; Coutelle C (1991) Biomed Biochem Acta 50(1):3-10). In general, the more closely matched two sequences are when hybridised together the more likely they are to resist denaturation. The denaturation brings about a marked difference in electrophoretic mobility and so the presence of sequence differences in samples can be determined by

comparing their behaviour during TGGE. The more dissimilar sequences will exhibit greater electrophoretic mobility.

Denaturing gradient electrophoresis (DGGE) is another method similar to TGGE except that the denaturing during electrophoresis is produced by a gradient of chemical denaturant in parallel with the voltage gradient (Coutelle C (1991) *supra;* Noll W W & Collins M (1987) PNAS 84(10):3339-3343).

Single stranded conformation polymorphism analysis (SSCP) is another known technique which relies on differing conformations adopted by closely related sequences on denaturation and rapid reannealing (Coutelle C (1991) *supra,* Sheffield V C (1993) Genomics 16(2):325-332; McGuire W L *et al.* (1991) Mol Endocrinol 5(11):1571-1577). Each conformation has a different electrophoretic mobility thus allowing them to be discriminated following electrophoresis.

The method known as chemical mismatch cleavage uses base differences between two otherwise identical sequences to produce real size differences in the sequences (McGuire W L *et al.* (1991) *supra).* The sequences under investigation are first hybridised to a control sequence and then chemicals which cleave at any mismatched bases are added. The approximate position of the mismatches can be determined by observing and analysing the altered electrophoretic mobility of the cleavage products.

The method of RNAse protection works in a similar manner to that of chemical mismatch cleavage except that the reference sequence is made from RNA and this is therefore sensitive to degradation by RNAse at the sites of any mismatches when hybridised to a target sequence (Osborne R L *et al.* (1991) Cancer Res 51(22): 6194-6198). Sequence variants are detected by looking for

the differences in electrophoretic mobility that they confer on the RNA sequence following RNAse digestion.

The method of enzymatic mutation detection works similarly to RNAse protection except that a DNA standard is used and mismatches are detected by cleavage with T4 endonuclease VII (Marshal R D *et al.* (1995) Nat Genet 9(2):177-183, Del Tito B J Jr (1998) Clin Chem 44(4):731-739; Youil R *et al.* (1995) PNAS 92(1):87-91). T4 endonuclease VII is a resolvase and its normal function is to cleave branched DNA intermediates that form during DNA replication. An additional ability of resolvases is to cleave DNA duplexes at sites that contain mispaired strands.

The indirect methods described above for detecting the presence of sequence variants all require electrophoresis or some other method of detecting mass of oligonucleotides, for example mass spectroscopy (Laken S *L et al.* (1998) Nat Biotechnol 16(13):1352-1356), or conformational differences between oligonucleotides. At best the indirect methods can only identify the approximate position of any base change. Moreover, they are not able to identify the nature of the change.

Methods of nucleic acid sequencing for example the Sanger Dideoxy Chain Terminator Method or the Maxam and Gilbert Chemical Cleavage Method are available and able to be used to detect the actual position and nature of any sequence variant. However, these methods are cumbersome suffering from the drawback of a low sample throughput. They are also methods which are expensive and laborious and require high resolution denaturing gel electrophoresis (Sanger F *et al.* (1977) PNAS 74(12):5463-5467; Maxam A M & Gilbert W (1977) PNAS 74(2):560-564.

Best T is a method which aims to improve the throughput of samples in sequencing by scanning for only the commonest changes which are usually C to T. Comparing the electrophoretic banding patterns produced by sample and control sequences when the x and y sequencing reactions are combined identifies the sequence positions at which alterations must have occurred.

Improved ways of sequencing many samples in parallel or rapidly in series and without using gel electrophoresis have been described. For example, WO 95/20053 (MRC) discloses a method of sequencing a nucleic acid, comprising either sequentially removing bases from the sequence of the nucleic acid a predetermined number at a time, with the product remaining from each step of predetermined base removal being ligated to a labelled adapter specific for said bases and including oligonucleotide sequence, or hybridising a primer to the nucleic acid to be sequenced and sequentially extending said primer a predetermined number of bases at a time, said added bases(s) being complementary to base(s) in the nucleic acid being sequenced, and each of said base addition steps being achieved by the use of a labelled adaptor specific for said bases and including oligonucleotide sequence containing said predetermine base(s); in either case, the label of said labelled adaptor being specific for its respective predetermined base(s).

Pyrosequencing (Ronaghi M *et al.* (1999) Anal Biochem 267(1):65-71) during which base additions are monitored by enzymatically converting the resultant pyrophosphate into substrates for luciferase thus producing detectable light avoids using electrophoresis but can only examine short stretches of sequence and has low discrimination.

Many of the methods described above have been used in combination with one other. It is possible, for example, to use the discriminating power of DNA polymerases with ligation assays. (For example, Abravaya K *et al.* (1995) Nucleic Acids Res 23(4): 675-682; Barany F (1991) PNAS 88(1):189-193; Wu D Y & Wallace R B (1989) Genomics 4(4): 560-569; Lizardi P M *et al.* (1998) Nat Genet 19(3):225-232).

Comparative nucleic acid sequence analysis can be performed by applying a sample to solid state arrays of oligonucleotide fragments (Gunderson K L *et al.* (1998) Genome Res 8(11): 1142-1153).

Demanding analytical situations arise when a wide variety of different mutations in a single gene each cause a single genetic disease, for example cystic fibrosis. For researchers it is difficult to anticipate where deleterious changes will occur, especially in a large gene. Screening therefore has to be restricted to known or common changes if the entire sequence is not to be examined. Examining genes for acquired alterations which may have clinical diagnostic significance is even more demanding. Cancer is a well known example. Cancerous cells are abnormal and as such are likely to have acquired significant sequence changes in their genes. Cancer cells are usually found in association with normal cells in the body. The proportions and types of cells in a cancer are likely to vary on a case by case basis. The sequence changes that are sought could comprise only a small proportion of the total making them more difficult to detect. Cancer cells themselves may also be heterogeneous for any given sequence change so that not all carrier cells may harbour the same change. This may further reduce the probability of a particular change being detectable. The small quantities of sample materials usually available may also adversely affect the likelihood of detecting certain sequence changes. Large amounts of material may not be

available from a human biopsy sample, especially in the case of cancer if it is at an early stage. There is sometimes insufficient material to detect any changes at all when using existing methods. This is particularly true if numerous possible changes are sought because they cannot always be sought at the same time in the same assay. Biopsies may therefore have to be divided between assays thus reducing the amount of material available to each. Also, samples are usually heterogeneous and this therefore necessitates that the methods of analysis have a high degree of discrimination. Problems with sample size or a desire to perform multiple tests means that analytical procedures need to be highly sensitive in order to detect changes. Existing methods do not necessarily provide the necessary sensitivity.

The inventor has identified a need for a high throughput method of sequence analysis which has the sensitivity to detect both the positions and identities of sequence changes right down to single base resolution and yet which is able to discriminate sequence variants when they exist as only a small proportion of the total sequences in a sample.

The inventor has also appreciated that any polynucleotide sequence can be broken up into what can notionally be regarded as a series of sequence "words" (i.e. fragments) of pre-defined length. For example, the sequence AAAA is made up of one 4 base word, or two possible 3 base words which both happen to be AAA, three possible 2 base words all of which are AA, and four possible one base words all of which are A. More complex sequences can similarly in theory be broken into sequence words and algorithms exist for this purpose (eg GCG - Staden). For example, the sequence ATTGCG has one 6 base word, two 5 base words ATTGC and TTGCG, three four base words ATTG, TTGC and TGCG, and so on for sequence words of reducing size. However, in some

complex sequences the number of apparent words may be less than the number of actual words of a given size because the same sequence words may occur at different positions in the overall sequence. For example, in the sequence ATTGCGCATTG (SEQ ID NO: 1), the 4 base word ATTG occurs twice. The theoretical collections of words derived from double stranded sequences may be even more complex since different words may be found in each strand of the sequence.

The inventor has also appreciated that pre-selected groupings of sequence words can be employed as a means of identifying features of interest in a sequence. For example, in two nearly identical polynucleotide sequences that differ only by a single base alteration, a number of different sequence words can be found to be different between the two sequences; the actual number of sequence words which differ depending on the length of the sequence words themselves. For example, in the sequence ATTGCGATTG (SEQ ID NO: 2) which differs by only a single base compared to a second sequence ATTGCCATTG (SEQ ID NO: 3), the three base words GCG, GAT found in the first sequence are GCC, and CAT in the second sequence. A way of finding out the identity of the base responsible for a single base difference between two sequences is to compare all of the three base differences between those two sequences, identify the words unique to each sequence, and then use them to deduce the sequences of nucleotides in the first and second sequences at or around the position in the sequence where the difference occurs. In the example above, the two sets of three overlapping 3 base words unique to each sequence can be aligned to show that there is a change from G to C in the middle of the 5 base word GCGAT.

Prior to the present invention, a problem in principle with comparing all possible words in lengthy polynucleotide sequences was the immensely large number of possible sequence words and their possible variations which need to be provided, hybridised to the polynucleotide and then analysed. The inventor has found that this problem can be solved by comparing groups of sequence words thereby reducing the number of comparisons that need to be made in an analysis. The groups of sequence words which need to be established for making comparisons are not random assortments of words, but groups of sequence words having a pre-determined relational integrity between one another. The relationships are such that the groups of words can be cross-referenced with one another in various ways when carrying out the process of identifying sequence word(s) associated with a given sequence difference and then deducing the originating sequences and the sequence difference. The individual sequence words do not have to be examined in isolation.

Hitherto, it has not been practicable or possible to produce a library of sufficient size and complexity, e.g. all possible sequence words of a given size, and all variations of those words, even by synthesising oligonucleotides in order to perform the kinds of analysis which the inventor has envisioned. The present inventor has surprisingly found ways of producing libraries of all possible sequence words and all of their variants. Moreover, the inventor has found ways of producing the sequence words as pre-defined groups so that they can be cross-referenced with one another in various ways so as to permit identification of differences between sequences of interest, particularly unknown differences between sequences.

According to a first aspect of the present invention there is provided a method of comparing first and second sample polynucleotides, comprising the steps of:

i) providing at least two different sub-populations of the first sample, each sub-population comprising a series of fragments of the first sample polynucleotide of known length, the 5' terminus of each fragment being located at a known position in the first sample polynucleotide;

ii) with each of the first sample sub-populations, providing a plurality of modification libraries by dividing the sub-population into a set of modification libraries and modifying the nucleic acid at a fixed position or plurality of fixed positions in each fragment, the modification libraries of the sub-populations between them providing for modification at each position in the first sample polynucleotide.

iii) contacting each modification library of each sub-population with the second sample polynucleotide under stringent hybridisation conditions and detecting the hybridisation of the second sample polynucleotide to the fragments of each modification library; and

iv) correlating the results of detection step (iii) to determine any differences between the first and second polynucleotides.

The first and second sample polynucleotides may be related to one another. In particular they may have at least 70% homology, for example at least 80, 90, 95 or 99% homology. They may for example be alleles of a gene.

The fragments of each sub-population may be of different lengths or they may be the same length.

Each modification library may have a different modification at said fixed position.

A series of fragments of known length may be obtainable from a sub-population or combination of sub-populations such that the sub-population or sub-populations form a contiguous series of fragments of the first sample polynucleotide. Alternatively, an overlapping series of fragments of the first sample polynucleotide may be provided.

The modification to each sub-population at the fixed position or fixed positions may be selected from the group of substitution, deletion and addition of a nucleotide, and inversion of a pair of nucleotides. The modification may be substitution and each sub-population being divided into twelve modification libraries, between them providing for each possible substitution of each nucleic acid, each modification library providing for one substitution of one nucleic acid. Alternatively, the modification may be substitution and each sub-population being divided into four modification libraries, between them providing for substitution by the same nucleic acid of each nucleic acid of the first sample polynucleotide.

The modification of the nucleic acids of a modification library may occur at the 3' or 5' terminus of the fragments.

A method according to the present invention may comprise prior to the step of contacting each modification library of each sub-population with the second sample polynucleotide, the additional step of labelling the fragments of each sub-population. The label may for example be selected from the group of a mass label, a chemical label, a ligand, an enzyme and a radiolabel. The label may be a chemical label comprising a coloured dye.

A method according to the present invention may comprise labelling the 5' or 3' terminus of the fragments of each sub-population.

Labelling may be performed when nucleic acids are modified to form the modification libraries.

The correlation step (iv) may comprise identifying any combinations of modification libraries which because of their character and relation to one another cannot give rise to any observed pattern difference, thereby allowing identification of any combinations of modification libraries actually responsible for hybridization (and non-hybridization) detected in step (iii), the respective characters and relationships of any modification library combinations thereby permitting identification of the nature and/or the position of the difference between the first and second sample polynucleotides.

The methods of the invention advantageously employ fewer steps than prior art methods. Moreover, the amount of sample processing prior to analysis is reduced and the subsequent detection of reaction products does not require specialised systems making it economic and therefore suitable for widespread and general use. Further advantages arise in that the methods can be tailored such that reaction products only arise in connection with sequence variations thereby reducing the need for any unnecessary analysis of the oligonucleotide fragments relating to the normal (reference) sequence. Also, multiple sequence positions can be analysed at the same time in a single solution.

In preferred embodiments the modification libraries are ones whereby substantially all fragments will be hydridisable under appropriately selected hybridisation conditions to a polynucleotide template under investigation (i.e. the second sample polynucleotide). The methods of the invention may therefore be designed so as to rely on a detection of any faults or failures in hybridisation between the fragments of the modification libraries and second sample

polynucleotide at any point along the polynucleotide sequence, for example. This is in contrast to other known methods which would look for an individual hybridisation event, e.g. between a single identifiable oligonucleotide fragment and the polynucleotide under investigation.

An ability to identify within a set of modification libraries separate sub-populations of oligonucleotide fragments which are related to each other in a pre-defined way and also to be able to cross-reference the sub-populations permits the sets of modification libraries of the invention to be used to "interrogate" the second sample polynucleotide. Usually, the process of interrogation would involve hybridisation of the sets of modification libraries to the second sample polynucleotide followed by the detection of some reaction, e.g. some fault, failure or change in hybridisation characteristics between oligonucleotides of one identifiable sub-population, and/or between oligonucleotides of two separately identifiable sub-populations. The sub-populations have "referential integrity" (i.e. meaning that sub-population members can always be identified as belonging to that sub-population and as a result they have a particular known sequence-related characteristic and that the individual sub populations can be identified as differing from one another in terms of at least one pre-determined sequence or structural characteristic). In practice this means that particular sequence information, usually general in character, is associated with the identifiable sub-population of fragments comprising the modification libraries. The pattern of reaction or hybridisation behaviour amongst a series of modification libraries exhibiting a variety of sequence-related characteristics can be established and then analysed using the method of the invention. By comparing the hybridisation behaviours and the sequence characteristics of the various modification libraries in various combinations the person skilled in the art will be able to deduce how precisely the sequence of the second sample polynucleotide

differs from the reference sequence (i.e. the first sample polynucleotide) which the subpopulations are derived from.

The sequence of the reference polynucleotide (the first sample polynucleotide) may be known or unknown. In circumstances where the sequence of the first sample polynucleotide is not known then the method of the invention still permits identification of the position and the identity of a mutation in the second sample polynucleotide.

Advantageously therefore the invention permits the location and identification of mutations without having to sequence polynucleotides and then compare sequences. Of course once the location and identity of a mutation is determined then the relevant part of the polynucleotide can be sequenced to obtain more detailed sequence information.

The contacting of the second sample polynucleotide with the modification libraries is preferably carried out under denaturing conditions although it would be possible to use non-denaturing conditions followed thereafter by denaturing conditions.

The sub-populations (and therefore the modification libraries) may be derived at least partially by fragmentation of the first sample polynucleotide but preferably entirely by fragmentation procedures. Thus the sub-populations may be derived from the first sample polynucleotide sequence at least in part by a method of oligonucleotide synthesis.

Sub-populations made by fragmentation of a selected polynucleotide sequence opens up advantageous possibilities over sub-populations produced in other

ways e.g. through chemical synthesis. For example, sub-populations obtained from a polynucleotide by degradation can comprise a number of oligonucleotide fragments exceeding a number which is presently practicable to produce by chemical synthesis. Thus, the number of possible sizes and sequence variations of fragments in a set of sub-populations and modification libraries is greatly increased relative to sub-populations/modification libraries made by chemical synthesis. The sub-populations of the invention may be characterised in that essentially they comprise a number of first sample polynucleotide fragments such that they are not produced by chemical synthesis.

A set of sub-populations (and their modification libraries) is also referred to as "a library".

A modification library will usually correlate with an ability to discriminate it from other modification libraries. The discrimination might be by way of observation relying as a marker or tag, e.g. a chemical tag, or simply through knowing how the particular modification library was produced by fragmentation, optionally including modification, from the original nucleic acid molecule. Similarly, sub-populations may be discriminated from one-another knowing how the particular sub-population was produced by fragmentation.

The particular composition of a library in terms of fragments, number of sub populations, number and kind of modification libraries, type of labelling of fragments, length of fragments, degree of overlap between fragments, etc is at the choice of the user and depends on the particular kinds of information which is needed to be known about a given polynucleotide sequence, whether the nucleotide sequence is known or not. In accordance with the methods of the invention the user may wish to "interrogate" a polynucleotide of interest in a

particular way and the library used in the method to perform the interrogation can be made to comprise the appropriate sub-populations of oligonucleotide fragments needed to perform that interrogation. The interrogation of the polynucleotide may proceed in a series of stages, each stage employing a different library and the design of each stage and its library may benefit from knowledge derived from the previous stage.

The methods of the invention therefore employ a complex library probe comprising families of oligonucleotides corresponding to the length of a polynucleotide of interest. The families of the complex probe are each characterised by reference to sequence or structural information but at a general level. Analysis of the behaviour patterns of the members of the complex probes on hybridisation to a target permit more specific sequence or structural information to be deduced. In other words, the methods of the invention permit the identification of very specific information from certain combinations of more general information.

Library sub-populations may be identifiable in that they are physically separate from each other and/or that each oligonucleotide fragment in a sub population is tagged with a tag specific to that sub-population.

When a tag is used it may be selected from a mass label, a chemical label, a ligand, an enzyme or a radiolabel. When a chemical label is used it may be selected from a coloured dye, preferably a fluorescent dye, optionally selected from TAMRA, FAM, ROX or JOE.

Modifications and variations of the oligonucleotide libraries of the invention are provided. At an extreme, libraries may comprise a multiplicity of sub populations

of oligonucleotides which in sum provide a comprehensive library of fragments in which every possible length of fragment is present, every possible position in the polynucleotide sequence maps with an end of a fragment, and every possible modified fragment of each of the above mentioned fragments is present, in the sense that fragments with every possible individual base substitution, insertion or deletion are present. Not all possible modifications of fragments need to be provided and the following is illustrative of the variety of modified sub-populations of library fragments which are possible and which may be selected by the use of the library depending on the particular circumstances of the polynucleotide sequence being analysed and the particular information which is being sought about the polynucleotide.

Modification libraries are derived by modification of existing library sub-populations.

The relationship of one sub-population / modification library with another is preferably characterised by the way or ways in which it was derived from the first sample polynucleotide.

The modification may be that the ends of the fragments in the sub-population map a fixed increment of bases from the respective ends of the reference polynucleotide.

The modification to the nucleic acid of the fragments of each sub-population at a fixed position or a plurality of fixed positions may be selected from:

(i) end base deletion of fragments;

(ii) end base substitution of fragments;

(iii) end base addition to fragments;

(iv) deletion of one or more bases a fixed number of bases from fragment ends;

(v) substitution of one or more bases a fixed number of bases from fragment ends;

(vi) insertion of one or more bases a fixed number of bases from fragment ends; and

(vii) inversion of a sequence of two or more bases, the inversion starting a fixed number of bases from fragment ends.

If a modification library is produced by making any of the changes (i) to (vii) noted above then in accordance with the invention a further modification library can be produced from part of the existing modification library by subjecting it to any other (i) to (vii) above.

Where there is an end base deletion, substitution or addition then this is preferably at the 3' end of the fragments although it may be at the 5' end of the fragments instead.

Where there is an end base substitution or addition then in preferred embodiments this yields a modification library of fragments all having the same end base selected from A, T, C, G or U. Alternatively, the modification may be such that an end base selected from A, T, C, G or U is substituted with a base other than itself selected from A, T, C, G or U.

Any modification to delete, substitute, insert or invert bases may start at a point up to and including position n/2, wherein n is the length of the fragment prior to modification.

In preferred methods of the invention, the contacted polynucleotide(s) and modification library fragments react by annealing under hybridizing conditions, preferably stringent hybridizing conditions, the reaction further comprising the step of carrying out ligation or polymerization.

The hybridisation conditions referred to are usually stringent although the degree of stringency may be adjustable by the user of the procedures of the invention in order to achieve a particular degree of hybridisation between modification library fragments and the second sample polynucleotide.

When the contacted second sample polynucleotide and modification library fragments react, a nuclease reaction (which may be an exo- or an endo-nuclease reaction) may take place, e.g. an endonuclease or cleavage reaction, preferably a DNase reaction, more preferably a ribonuclease reaction. When the reaction is a cleavage reaction then a suitable enzyme is exemplified by T4 endonuclease VII. Alternatively, the reaction may be manifest as the avoidance or promotion of a chemical event or modification.

The products of any reaction may be identified and/or analysed by a method of determining a parameter selected from size, mass, charge, length or ligand binding. The method of identification and/or analysis may be selected from electrophoretic procedures, preferably gel electrophoresis; mass spectrometry; chromatographic procedures, preferably high pressure liquid chromatography (HPLC); fast protein liquid chromatography (FPLC), liquid chromatography, TLC or GLC. The products of any reaction may be tested for the presence or absence of labels.

Also provided according to the present invention is a method according to the first aspect of the present invention, the step of contacting each modification library of each sub-population with the second sample polynucleotide being carried out in the presence of the sub-population.

In such a method, the fragments of each modification library may be modified such that they have a 3' dideoxynucleotide, each modification library sub-population being labelled at the 5' terminus, the step of contacting each modification library of each sub-population with the second sample polynucleotide being performed in the presence of a ligase enzyme, and the detection of hybridisation comprising detecting ligation products.

Also provided according to the present invention is a set of modification libraries of at least two different sub-populations of a sample polynucleotide, each sub-population comprising a series of fragments of the sample polynucleotide of equal length, the 5' terminus of each fragment being located at a position in the sample polynucleotide n nucleotides from each adjacent fragment wherein n is at least 2, the modification libraries comprising sub-divisions of each sub-population having a modified nucleic acid at a fixed position or plurality of fixed positions in each fragment, each modification library having a different modification at said fixed position, and the modification libraries of the sub-populations between them providing for modification at each position in the first sample polynucleotide.

In such a set of modification libraries, the fragments of each sub-population may be of the same length.

In such a set of modification libraries, at least one sub-population may comprise a series of fragments of length n such that the population forms a contiguous series of fragments of the first sample polynucleotide.

In such a set of modification libraries, the modification to each sub-population at the fixed position or fixed positions may be selected from the group of substitution, deletion and addition of a nucleotide. With the modification being substitution and each sub-population being divided into twelve modification libraries, between them providing for each possible substitution of each nucleic acid, each modification library provides for one substitution of one nucleic acid. With the modification being substitution and each sub-population being divided into four modification libraries, between them they provide for substitution by the same nucleic acid of each nucleic acid of the first sample polynucleotide.

In such a set of modification libraries, the modification of the nucleic acids of a modification library may be at the 3' terminus of the fragments.

In such a set of modification libraries, the fragments of each sub-population may be labelled. The label may be selected from the group of a mass label, a chemical label, a ligand, an enzyme and a radiolabel. The label may be a chemical label comprising a coloured dye. The label may be at the 5' terminus of the fragments of each sub-population.

Also provided according to the present invention is the use of a set of such modification libraries in a method according to the first aspect of the present invention.

Substantially all of the different modification libraries are preferably hybridisable to the sample polynucleotide, including sequence variants of that polynucleotide, preferably under stringent hybridizing conditions. However, the stringency of the hybridizing condition may be adjusted in order to achieve the functional effect of substantial hybridization by all modification libraries.

Modification library fragments will preferably overlap in sequence with other modification library fragments. In a particularly preferred embodiment the modification libraries comprise substantially all possible overlapping oligonucleotide fragments derived from the polynucleotide.

In other embodiments modification library fragments overlap in sequence with other modification library fragments by n-m nucleotide residues, wherein n is the length of the fragments and m is in the range 2 to (n-3).

In yet further embodiments each modification library fragment overlaps in sequence with at least 2x(n-2) other fragments in the modification library, wherein n is the fragment length .

In still further embodiments modification library fragments overlap in sequence with other fragments by n-1 nucleotide residues, wherein n is the length of the fragments.

The fragments of sub-populations and modification libraries may be of uniform length or of different lengths.

In particularly preferred embodiments the sub-populations may comprise fragments corresponding to substantially all possible n mers of the first sample

polynucleotide, wherein n is in the range 3 to z, wherein z is the length of the first sample polynucleotide minus 1.

When bases are added to fragments in order to provide modification libraries, the fragments are associated with the label so that the position and/or identity of one or more additional bases in a given fragment can be recognized, optionally wherein:

a) the labelling is direct such that the labelled base is the added base; or

b) the labelling is indirect such that it signifies the position and/or identity of one or more additional bases in the fragment.

The additional base may serve to prevent ligation or chain extension therefrom, in which case the additional base is preferably a dideoxynucleotide.

When there are modification libraries arising through substitutions then the fragments of the modification libraries are labelled in such a way that the identity of a particular base substitution can be recognized in a given fragment. When there are modification libraries arising through deletions then the fragments of modification libraries are labelled in such a way that the identity of a particular deletion can be recognized in a given fragment. When there are modification libraries arising through insertions then the fragments of modification libraries are labelled in such a way that the identity of a particular insertion can be recognized in a given fragment.

In a further aspect the invention provides a process for preparing a library as hereinbefore described comprising the fragmentation of a polynucleotide.

Sub-populations are preferably identifiable from each other by their being kept physically separated from one another, and modification libraries are preferably identifiable by individual fragment members being tagged specifically to indicate the modification which took place to provide the modification library of which they are part.

The fragmenting of the polynucleotide first preferably provides a population of oligonucleotides of non-uniform length but having coterminal ends, the fragmenting process then preferably comprising the cleaving of said non uniform length oligonucleotides a predetermined number of bases from their non-coterminal ends, optionally removing the resulting population of oligonucleotide fragments without coterminal ends from the remaining population of oligonucleotides of non-uniform length and coterminal ends.

The initial fragmentation of the polynucleotide may be by sonication. Coterminal ends of the oligonucleotides of non-uniform length are preferably generated by fragmenting the polynucleotide with a nuclease enzyme, preferably an endonuclease.

When the oligonucleotides of non-uniform length are cleaved then this preferably comprises ligating a first adaptor to the non-coterminal ends of the oligonucleotides of non-uniform length, said adaptor including a recognition site for a first nuclease having its recognition site displaced from its cleavage site by a number of bases sufficient to cause cleavage by said predetermined number of bases from the non-coterminal ends, and reacting the oligonucleotides of non-uniform length having the first adaptor linked thereto with said first nuclease.

The cycle of ligation of a first adaptor and reaction with a first nuclease may be repeated at least one further time, preferably a number of times.

The process of the invention preferably further comprises the step of removing the first adaptor from the cleaved oligonucleotides.

The adaptored fragments are preferably ligated to a second adaptor, said adaptor including a recognition site for a second nuclease having its recognition site displaced from its cleavage site by a number of bases sufficient to cause cleavage by said predetermined number of bases and the ligation product is reacted with a second endonuclease. The second adaptor and second endonuclease may be the same as the first adaptor and first endonuclease. Alternatively, the second adaptor may be ligated to the non adaptored ends of the adaptored fragments.

The second adaptor may be ligated to the first adaptor of the adaptored fragments such that on reaction with the second nuclease both the first and second adaptors are cleaved from the said oligonucleotides.

Thus sub-populations of a sample polynucleotide may be readily produced, the sub-populations comprising a series of fragments of known length, the 5' terminus of each fragment begin located at a known position in the sample polynucleotide.

In a further aspect, the invention provides a method of producing a library as hereinbefore described comprising sequential removal of oligonucleotides of n bases from the end of a polynucleotide sequence, thereby providing sub populations of said polynucleotide sequence wherein n is in the range 3 to 50,

preferably 3 to 22, more preferably 3 to 16. Prior to sequential removal of oligonucleotides of n bases, the intact starting polynucleotide sequence may have one, two, three or more nucleotide residues removed from one end thereof, thereby resulting in separate sub-populations of oligonucleotides of n bases whose ends are characterised by being related to the intact sequence by mapping at or by a multiple of n bases from positions 1, 2, 3, 4 or more respectively.

Removal of oligonucleotides of n bases is preferably effected by ligation of an adaptor to the polynucleotide sequence, the adaptor having a recognition site for a nuclease enzyme capable of cleaving nucleic acid at a site a pre determined distance from the recognition site, followed by reacting the adaptored polynucleotide with the nuclease enzyme. Preferred adaptors have a recognition site for a Type II restriction enzyme and therefore the preferred nuclease enzyme is a Type II restriction enzyme.

The cycle of ligating the adaptor and reacting with nuclease enzyme may be repeated more than once, preferably a multiplicity of times.

In a further aspect the invention provides a kit for producing a library as hereinbefore described comprising one or more of a nuclease enzyme capable of cleaving nucleic acid at a site a pre-determined distance from a recognition site, at least one oligonucleotide adaptor having a sequence comprising the recognition site for the said nuclease, a ligase enzyme, an oligonucleotide adaptor and means for labelling the same and a vector cloning kit for use in the elimination of recognition sites for the said nuclease from a polynucleotide sequence.

The invention provides for the use of a library as hereinbefore described in the methods of the invention.

The invention also includes the use of a sequence ladder as an oligonucleotide library in any of the methods of the invention as hereinbefore described.

The invention further includes apparatus, e.g. a computer comprising means to perform the correlation step of the method as hereinbefore described. A computer may therefore be loaded with a program which performs the correlation and thus provides a way of automating performance of at least some of the steps of the methods of the invention. Within the scope of the invention is apparatus arranged and configured to perform the entirety of the methods.

Preferred embodiments of the invention will now be described both more generally and then by way of specific examples having regard to the drawings in which:

Figure 1 shows in principle how a library of the invention is produced by cyclic removal of 4 base fragments from the end of a sequence (SEQ ID NO: 40) of interest. (a) shows the sequence of interest; (b) shows the position of cleavage 4 bases from the end; (c) shows that the first cleavage releases 4 bases from the fragment end; (d) shows that the second cleavage releases a further 4 bases from the fragment end; and (e) shows that the third cleavage releases a further 4 bases from the fragment end.

Figure 2 is like figure 1 except that 5 base fragments are removed, and shows the production of a library by cyclical removal of 5 base fragments from the end of a sequence of interest (SEQ ID NO: 40). (a) shows the

sequence of interest; (b) shows the position of cleavage 5 bases from the end; (c) shows that the first cleavage releases 5 bases from the fragment end; (d) shows that the second cleavage releases a further 5 bases from the fragment end; and (e) shows that the third cleavage releases a further 5 bases from the fragment end;

Figure 3 shows the way in which adaptors can be used to remove a desired number of bases from the end of a polynucleotide sequence (part of SEQ ID NO: 40) of interest. (a) shows the Type11s site situated to cut at the end of the adaptor; (b), (e), (h) and (k) show ligation; (c), (f), (i) and (l) show cutting using the Type11s restriction endonuclease; (d) shows the Type11s site situated to cut 1 base after the end of the adaptor; (g) shows the Type11s site situated to cut 2 bases after the end of the adaptor; and (j) shows the Type11s site situated to cut 3 bases after the end of the adaptor;

Figure 4 shows a reaction scheme for production of a library comprising end base removal from a polynucleotide of interest (SEQ ID NO: 40) followed by cyclic ligation and cutting of the resultant polynucleotide to remove 6 base fragment s from 0, 1 and 2 bases from the end of the sequence of interest. (a) shows the target sequence. End base removal pretreatment then takes place followed by end base removal of 1, 2 and 3 bases ((b), (c) and (d) respectively); (e) shows the first cycle of fragment production for the library; and (f) shows the products of a first cycle of fragment production for the library;

Figure 5 shows the start of a process of fragmenting a polynucleotide using a blunt ended adaptor containing the site for a Type11s restriction endonuclease which leaves a 2 base 3' overhang, and a Type IIs restriction endonuclease. (a) shows the nucleic acid of interest as a general

sequence (SEQ ID NO: 41); (b) shows ligation of the adapter; (c) shows the ligated adapter and sequence of interest (SEQ ID NO: 42); and (d) shows cutting using the Type11s restriction endonuclease;

Figure 6 shows the next step and indicating the subsequent steps in the process of Figure 5. Adaptors with a 2 base 3' overhang are used. (a) shows the nucleic acid of interest (a fragment of SEQ ID NO: 42); (b) shows ligation of the adapter; (c) shows the ligated adapter and sequence of interest; and (d) shows cutting using the Type11s restriction endonuclease;

Figure 7 shows a process in which adaptored fragments have their adaptors removed by the use of a second adapter with the site for a Type11s restriction endonuclease that leaves a 2 base 3' overhang. (a) shows the adaptered first fragment; (b) shows the second adapter and its ligation; (c) shows the ligation product; and (d) shows cutting using the Type11s restriction endonuclease;

Figure 8 shows an alternative process to that of Figure 7 in which adaptored fragments have their adaptors removed. (a) shows the adaptered first fragment; (b) shows the second adapter and its ligation; (c) shows the ligation product (SEQ ID NO: 43); and (d) shows cutting using the Type11s restriction endonuclease;

Figure 9 shows how adaptors can be used to create a library population in which A residues are removed from the penultimate 3' end. (a) shows a member of a fragment library (i.e. a fragment of a sub-population); (b) shows the adapter and its ligation; (c) shows the ligation product; and (d) shows

cutting using the Type11s restriction endonuclease to produce a modified fragment;

Figure 10 shows how further adaptors can be used to complete a process in which a base is substituted in one strand and provided with a complementary base in the other strand of a library fragment - C's are used to replace T's previously found opposite to the penultimate A's at the chosen 3' end of fragment library members. (a) shows a member of a fragment library (i.e. a fragment of a sub-population); (b) shows the adapter and its ligation; (c) shows the ligation product; and (d) shows cutting using the Type11s restriction endonuclease to produce a modified fragment;

Figure 11 shows how appropriately constructed Type IIs adaptors and enzyme cutting are used to modify library fragments so that a 3' overhang is converted into a 5' overhang for labelling 9 bases from a change. (a) shows a sequence to be labelled and an adapter; (b) shows an N to C change at the 5' end of the sense strand of the sequence being labelled (fragment of SEQ ID NOs: 44 and 45), the adapter and their ligation; (c) shows the ligation product of (b) (sense strand is SEQ ID NO: 44, anti-sense strand is SEQ ID NO: 45); and (d) shows cutting with the Type11s restriction endonuclease;

Figure 12 shows how the 5' overhang fragments of Figure 11 are 3' end labelled with fluorescently tagged dideoxy-nucleotides 9 bases from the change. (a) shows a generic sequence to be labelled; (b) shows the sequence of Figure 11(d); (c) shows the extension of the 3' end of the sense strand by the activity of a DNA polymerase and labelled dideoxy terminated nucleotides; and (d) shows the product of polymerisation, having a labelled 3' A base on the sense strand 9 bases from the substitution;

Figures 13 and 14 show a similar process to that shown in Figures 11 and 12 except that the Type IIs adaptors are constructed so as to achieve a labelling of fragments in a way which labels the identity of a base change made 10 residues away from the label - the conversion of a 3' overhang to a 5' overhang for labelling 10 bases from the change. (a) shows a sequence to be labelled and an adapter; (b) shows an N to C change at the 5' end of the sense strand of the sequence being labelled (fragment of SEQ ID NOs: 44 and 45), the adapter and their ligation; (c) shows the ligation product of (b) (sense strand is SEQ ID NO: 44, anti-sense strand is SEQ ID NO: 45); and (d) shows cutting with the Type11s restriction endonuclease;

Figure 14 shows how the 5' overhang fragments of Figure 13 are 3' end labelled with fluorescently tagged dideoxy-nucleotides 10 bases from the change. (a) shows a generic sequence to be labelled; (b) shows the sequence of Figure 13(d); (c) shows the extension of the 3' end of the sense strand by the activity of a DNA polymerase and labelled dideoxy terminated nucleotides; and (d) shows the product of polymerisation, having a labelled 3' T base on the sense strand 10 bases from the substitution;

Figure 15 shows how an oligonucleotide can be labelled primarily to identify the penultimate 3' base species and secondarily the position of that 3' base relative to a known feature of the sequence - in this case the addition of a base specific 3' label depending on the penultimate base in a 3' overhang. (a) shows a sequence to be labelled and an adapter; (b) shows an N to C change at the 5' end of the sense strand of the sequence being labelled (fragment of SEQ ID NOs: 44 and 45), the adapters (having blue, cyan, green and red labels) and their ligation; and (c) shows the ligation product of (b) (sense strand is SEQ

ID NO: 44, anti-sense strand is SEQ ID NO: 45), ligation only having taken place with the adapter having the green label;

Figure 16 shows generally how template directed ligation will only occur between oligonucleotides which anneal in juxtaposition along a template with no mismatch. (a) is denaturation; (b) is annealing; and (c) is ligation;

Figure 17 shows how a mutation introduced into the template of Figure 16 permits ligation to take place between oligonucleotides which would not otherwise have been ligatable. (a) is mutation; (b) is denaturation; (c) is annealing; and (d) is ligation;

Figure 18 corresponds to Figure 16 except that actual base identities (SEQ ID NO: 46) are given rather than being a schematic Figure;

Figure 19 corresponds to Figures 16 and 18 except that the template (SEQ ID NO: 47) has a point mutation. (a) is mutation; (b) is denaturation; (c) is annealing; and (d) is ligation;

Figure 20 shows an example of a short arbitrary sequence (SEQ ID NO: 46) and the deletion libraries that can be produced from it. (a) is 6 base fragments produced cyclically from position 1 and 1 base end deletions; (b) is 7 base fragments produced cyclically from position 1 and 2 base end deletions; (c) is 6 base fragments produced cyclically from position 2 and 1 base end deletions; and (d) is 6 base fragments produced cyclically from position 2 and 2 base end deletions;

Figure 21 shows how an end base deleted fragment library can be used to detect a deletion in a polynucleotide. 10 is the target. 20 is the site of target deletion. 30 are end deletion fragments. 40 is the deleted target. 50 is the deletion. 60 is the ligation product;

Figure 22 shows a table of examples of the possible additions and insertions to fragments of an arbitrary sequence (SEQ ID NO: 46);

Figure 23 shows how end base addition libraries can be used to detect an insertion. Target 70 has insertion site 80. Prior to insertion, end addition fragments 90 bind as shown at (a). In inserted target 100, end addition fragments 90 bind as shown at (b) and can be ligated as shown at (c).

Figure 24 shows how information from fragment libraries can be combined to interrogate a multiplicity of samples which differ from one another but only slightly. Sequences numbered 1-8 are SEQ ID NOs: 48-55; and

Figure 25 shows the annealing and ligation of specific oligonucleotides (SEQ ID NOs: 57,58) to a template (SEQ ID NO: 56) as described in Example 1.

In each of the drawings and in the following description, n represents by convention any base selected from a T C or G. Thus, sequence portions or sequences denoted by nnnnnn.. etc represent populations of sequences comprising up to 4^x individual sequence members, wherein x is the number of nucleotides in the sequence.

In what follows are generalised descriptions of sub-populations and modification libraries of the invention, how to produce them and how to use them in methods of determining the position and identity of a difference between two nucleotide sequences (i.e. first and second sample polynucleotides), in particular any single base variation between the sequences.

One of the parameters to be decided is the sequence word length used for the analysis. Usually, words are chosen so as to be sufficiently long to be specific in the chosen conditions of hybridisation to one sequences of interest but not so long that they cross react with the other sequence it is being compared with. Sequence words of about 17 bases in length (17 mers) have been found to be particularly useful.

Liquid libraries with relational integrity

The invention provides and makes use of sets of sub-populations of a sample polynucleotide comprising oligonucleotide fragments of a pre-determined length or lengths. The sub-populations are produced directly from the sequence of interest. If the sequence of interest is non-linear, e.g. in the form of a plasmid, then the sequence is linearised first. The sequence is degraded in a sequential and uniform manner by a fixed number of bases at a time and this number corresponds to the sequence word length which has been chosen. Figure 1 illustrates the process wherein sequence words of four bases are produced. Figure 2 illustrates the same thing for sequence words of five bases. The degradation is carried out by using a restriction enzyme and one or more adaptors which permit the restriction enzyme to act on the sequence of interest.

Generally, adaptors are reagents that can be added to a nucleic acid for the purpose of achieving further modifications to that nucleic acid. In this case, the adaptor is a nucleic acid that can be ligated to one end of the sequence of interest (the sample polynucleotide). The adaptor provides the means to degrade the sequence of interest a predetermined number of bases at a time from the end to which it is attached. In this example the adaptor has a site for a Type IIs restriction endonuclease and further sequences of nucleotide residues. The endonuclease site is located in relation to the further sequences so that restriction cutting occurs a desired, predetermined number of bases downstream. Type IIs restriction endonucleases are restriction endonucleases that recognise a specific nucleotide sequence but cleave nucleotides a fixed number of bases from the recognition sequence. Adding the restriction endonuclease to the products of a ligation reaction between a Type IIs adaptor and the sequence of interest therefore results in cleavage of the sequence of interest the predetermined number of bases from its end. Repetition of this process of ligation and cutting results in an entire sequence of interest being broken up into a population of fragments of predetermined size, thus providing a sub-population of the sequence of interest.

The above sub-population does not contain every possible sequence word of predetermined size but only sequence words found at positions corresponding to the predetermined increment of bases from the end of the initial sequence. This introduces a relational integrity into the sub-population.

a further sub-population of fragments can be made by removing a single nucleotide from the end of the sequence of interest prior to the cycles of ligation and cutting as already described. This introduces relational integrity into a library comprising the sub-populations and is achieved by using another kind of adaptor

in which the recognition site for the Type IIs restriction endonuclease is located in the adaptor so that when ligated to the target sequence exposure to the Type IIs restriction endonuclease results in a cleavage one base into the target sequence. This pre-cleaved sequence is then used as the starting material for the cyclic ligation and cutting process described previously and this results in a sub-population as before except that it contains different sequence words all of uniform length each related by the fact that they are derived from a starting point which is one base in from the start of the sequence of interest. In other words the 5' ends of the fragments map to position 2 or a position which is a multiple of n bases therefrom, wherein n is the fragment length (ie position 2).

Adaptors can be designed so that the starting point for cyclic ligation and cutting can be from any desired point from the end of the sequence of interest right up to the length of the intended sequence words to be produced. Figures 3 and 4 illustrate examples of this process if a separate library sub population is made from each possible starting point then in sum the library covers all possible sequence words from the original sequence. If the sub populations can be distinguished from one another, for example by keeping them separate from one another and employing them separately or in preselected combinations, then each word in an individual library has effectively been labelled with regard to its possible distance from the end of the original sequence.

Library combinations

Many combinations of sub-populations can be envisioned. Appropriate positioning of the Type IIs recognition site in adaptors used in cyclic ligation and cutting degradation can allow sub-populations of other sequence word lengths to be produced. In general, positioning of the recognition site in the adaptor so

that on ligation it lies closer to the sequence of interest will result in longer sequence words, whilst positioning of the recognition site in the adaptor so that on ligation it lies further from the sequence of interest will result in shorter sequence words.

Sub-populations of mixed sequence word length may be produced by employing two or more adaptors.

Sub-populations may also be produced by combining the adaptors which are used for an initial removal of bases from one end of the sequence into one or more pools. Families of fragments can thereby be produced covering a range of single base deletions from one end of the original sequence and up to the intended sequence word length. Cyclic endonuclease degradation of this material will still result in all possible words in the range of sub-populations used. Combining all of the above mentioned adaptors together will produce all of the possible words in a single sub-population.

The various adaptors and sub-populations can be used and produced individually such that sub-populations are distinct from one-another but that their use as individual (distinct) members of a library will provide the library with all of the possible words.

In the possibilities described above the cyclic endonuclease degradation of the sequence of interest may be carried out as a continuous process. However, this need not always be the case because due to purification or inactivation of the enzymes used after each ligation and each cleavage step the degradation can be limited to just a particular number of cycles. Sub-populations of sequences

deleted to precise distances from the end of a sequence of interest can therefore be provided.

a sequence to be divided into sequence words to form a sub-population may, because of the ligation step, remain covalently attached to the adaptor following the cutting step. Sequence words produced by the process still attached to adaptors may have had those adaptors removed. The first adaptor can be removed by ligating a second adaptor to the first adaptored sequence words. The second adaptor has an additional recognition site or sites for a Type IIs restriction endonuclease which can be the same as or different to the first. The site(s) are situated so that the first and second adaptors can be removed from the sequence words by cleavage with the restriction endonuclease. When double stranded sequences are used, it is also possible to arrange for either the adaptors or the sequence of interest, but not both, to lack a 5' terminal phosphate. When joining is catalysed by ligase this results in one covalently joined strand and one strand that is not joined. The latter is said to contain a nick. It can be arranged that this nick results in the sequence words of interest not being joined to the adaptor used.

The identification of single base sequence differences between sequences is made possible by the invention, both in terms of the nature and position of the difference. Sets of sub-populations corresponding to all possible differences can be produced in accordance with the invention from the sub-populations of sequence words described above. By maintaining separately identifiable library subpopulations the referential integrity between the sequence words and their altered modification library counterparts can be maintained. Another way of introducing referential (relational) integrity between library sub-populations is to use chemical forms of labelling as will be described in more detail later.

Production of all possible sequence changes with respect to the ends of the sequence words in a sub-population is illustrated in Figure 5 to 10. Figures 5 and 6 show how in principle a polynucleotide can be fragmented with adaptors and a Type IIs restriction endonuclease in the way previous described except that the adaptors are constructed so that the resulting adaptored sequence words have a 2 base 3' overhang.

Figure 5 shows how a blunt ended adaptor is used to start the fragmentation process and Figure 6 shows continuation of the fragmentation using adaptors with a 3' 2 base overhang.

The single adaptor represented in Figure 6 is just one of a population of adaptors which are required and therefore constructed in order to fragment the entire polynucleotide sequence into adaptored fragments having a 2 base 3' overhang.

Figure 7 shows how a population of second Type IIs adaptors can be ligated to the first adaptored fragments and then reacted with Type IIs endonuclease to yield sequence fragments free of adaptors. In this instance the restriction cuts are made in the same orientation, one further downstream from the other.

Figure 8 shows an alternative scheme for release of adaptors from sequence fragments. Again, a population of second Type IIs adaptors is ligated to the first adaptored fragments, but, in contrast to the scheme of Figure 7, to the non-adaptored ends of the fragments. The second adaptors are constructed so that on reacting the doubly adaptored fragments with Type IIs endonuclease, restriction enzyme cutting occurs downstream from the adaptors.

The resultant sub-population of fragments is divided into four equal samples. The fragments of each sample are ligated separately to an adaptor which is specific in its recognition of one of the four possible bases at the end of the sequence words. Thus, for each of the four starting samples, four modification libraries are generated, one for each of the four bases giving a total of 4 x 4 = 16 separate modification libraries. The adaptors in the populations of adaptors used in the process each have the recognition site for the Type IIs restriction endonuclease that was used to produce the initial population of sequence words. The recognition site is situated in the adaptors so that the base on the end of the fragments (whether a T C or G) recognised by the relevant adaptor in the population of adaptors is (on exposure to endonuclease) removed by digestion. Further adaptors are then added to each modification library of fragments so that when ligated they cause replacement of the removed base with one (or more if desired) of the three other possible alternative bases. These further adaptors also contain a site for the Type IIs restriction endonuclease and it is situated so that after ligation, removal of the further adaptor by cleavage with the restriction endonuclease leaves behind the fragments comprising the alternative base(s) at their end.

Figure 9 shows how appropriately constructed Type IIs adaptors can be used to remove penultimate a residues found at the 3' end of fragments.

Figure 10 shows how other appropriately constructed Type IIs adaptors can be used to further modify the fragment products of Figure 9. Overall, the result is a C residue added to the 5' → 3' strand and the penultimate a residue at the 3' end of the 3' → 5' strand is replaced with a complementary G residue so that it is no longer the penultimate base but the third base from the 3' end.

a set of modification libraries with referential (relational) integrity can be produced from each possible alternative base at given position(s), or all alternative bases can be substituted in a single modification library.

The modification libraries of the invention do not just include populations of fragments in which bases have been replaced at the ends of the fragments. Adaptors with complementary ends that result in an exchange of bases further into the sequence words can readily be made. In general, the longer the cohesive ends in an adaptor the more efficient the adaptor is at distinguishing sequences complementary to those cohesive ends. Also, positioning of the restriction endonuclease recognition site within the adaptor will determine the extent of cleavage into the target sequence and the positioning can also be arranged so that cleavage occurs so as to include at least part of the cohesive end of the adaptor.

Modification libraries containing sequence words with deletions, additions or insertions

In a similar way to that described above, modification libraries of fragments comprising deletions, additions or insertions of bases at the ends of the sequence words can also be produced. In the case of deletions at the ends of sequence words, the bases that have been specifically removed are simply not replaced. Successive deletions are possible to produce yet more modification libraries. Internal deletions are achieved by removing multiple bases and adding back less than the number removed.

Additions are achieved by adding adaptors to the sequence words and arranging the positioning of the Type IIs restriction endonuclease site within the adaptor so

that it leaves behind the desired additional bases. These additional bases can either be added to cohesive ends having particular sequences or they can be added to all ends.

Insertions can be achieved by cleaving within the original cohesive end and adding bases which result in the desired insertion at the site of cleavage and renewal of the remaining bases of the cohesive end once the adaptors used for the modification have been removed. All combinations of modification are possible, for example, deletion and addition.

Modification libraries containing modified oligonucleotides

Modification libraries with other useful referential integrity properties can be provided. For example, labelling of sub-population fragments can be carried out in a base specific manner at each end of a sequence word. This could be used as shown in Figures 11 to 15, to separate out individual members of a sub-population into groups according to the bases at their ends. Use of a Type IIs restriction endonuclease which produces a 5' overhang as shown in figures 11 and 12 or figures 13 and 14, for example, allows the ends of members of a modification library to be labelled in a base specific manner. In the case of what is shown in figures 11 and 12 or figures 13 and 14, advantage is taken of the availability of the four dideoxynucleotides terminators ddATP, ddCTP, ddGTP and ddTTP, each of which are labelled with a different fluorescent dye having a different emission wavelength. a DNA polymerase, for example Taq DNA polymerase, is used to add these dideoxy nucleotides in a template dependent manner to the 3' ends of sequence words in a modification library. This requires that the 3' ends are recessed or can become recessed. The dideoxy nucleotides are referred to as terminators because once added to the nucleic acid chain then

their 3' hydrogen (in place of the usual terminal hydroxyl group) prevents further additions of nucleotides. The DNA polymerase adds the dideoxy terminators to the corresponding recessed 3' ends according to Watson and Crick base pairing rules, ie dideoxy adenosine triphosphate with thymidine monophosphate, dideoxy cytosine triphosphate with deoxyguanidine monophosphate, dideoxy guanosine triphosphate with deoxycytodine monophosphate and dideoxy thymidine triphosphate with adenosine monophosphate.

Adaptors carrying a label that is specific to the base(s) found at their end can similarly be ligated to the sequence words to label those sequence words according to their end bases as shown in Figure 15. In this case it is not important whether the adaptors have 5' or 3' overhangs, only that the overhangs correspond in length and type to those found on the sequence words. In Figure 15 sequence words with 2 base 3' overhangs are labelled according to the bases found at their 3' end by ligation to corresponding adaptors which have been labelled according to the ends with which they are complementary. Figure 15 shows how an oligonucleotide can be labelled according to firstly the identity of the penultimate 3' base and secondly the position of that base relative to a known feature. In this case labelling is through the use of adaptors. The adaptors are labelled with a different dye according to the base with which they are complementary on their 3' end. Ligation and therefore labelling is only possible between oligonucleotides with complementary ends thus label is incorporated according to the 3' base which is present. The known feature is the base of a known number of base positions away at the opposite end of the oligonucleotide to be labelled. In this case an N to C change at the 5' end 11 bases from the labelled position.

The sequence of interest may itself include sites for the Type IIs restriction endonuclease and these sites could bias the representation of certain sequence words in the libraries. If it arises, this problem can be avoided by treating the polynucleotide of interest first with the relevant endonuclease and then subcloning the resulting fragments by methods well known in the art so that sites for the Type IIs restriction endonuclease used do not occur within the sequences of interest used to make libraries.

Relational integrity

The power of libraries of sub-populations and their modification libraries having relational integrity can be illustrated by the example of detecting any possible base difference between two otherwise identical sequences. Four sub-populations comprising all possible words of a predetermined length, wherein the predetermined word length is different in each sub-population, are prepared from the sequence of interest. Each sub-population is made by sequential fragmentation (by cyclical ligation and cutting) of the starting polynucleotide. The population of starting polynucleotides has been treated so that some polynucleotide molecules have one or more end bases removed thereby providing different start points for the sequential fragmentation. The start points cover the range of bases up to the length of the chosen predetermined length of the sequence words of the sub-population. Each of the sequence words in the sub-population is then used to form a modification library by labelling it at its 3' end with a fluorescently labelled dideoxy nucleotide in such a way that the label identifies the base that would normally be found at that position in the original sequence. These modification libraries will be referred to as labelled 3' end base modification libraries.

a further sub-population comprising all possible words of a predetermined length is made from the original sequence as described above. This sub-population is then divided into 12 equal samples. All possible end base replacement reactions are performed on these sub-populations, one per sub-population. In a first modification library all 3' deoxy adenosines are replaced by 3' deoxy cytosine, in a second modification library all 3' deoxy adenosines are replaced by deoxy guanosine and so on, i.e. A to C, A to G, A to T, C to A, C to G, C to T, G to A, G to C, G to T, T to A, T to C and T to G. The 5' ends of the oligonucleotides in all of these modification libraries is labelled in a way which allows them to be identified or isolated. Examples of labels include a fluorescent dye, a mass label, biotin or a hapten, for example, digoxygenein. These modification libraries are referred to herein as end base replacement modification libraries.

Each possible labelled 3' end base modification library is then mixed with each possible end base replacement modification library, ie two modification libraries per mixture making 48 mixtures in all. A sequence of interest (template) is then added to each of the mixtures to examine it for the presence of a suspected and unknown single base sequence difference compared to the sequence from which the modification libraries were made. The added sequence is hybridised to the oligonucleotides in each of the mixtures. The hybridisation conditions used are such that oligonucleotides capable of hybridising to the template in juxtaposition will do so. A ligase is then added so that any hybridised and juxtaposed oligonucleotides can be ligated together. Ligase has a certain fidelity so that reaction conditions can be chosen so that ligation will only exceptionally occur between hybridised and juxtaposed oligonucleotides exhibiting a base mismatch at the juxtaposed ends. Thus ligation conditions are selected so that effectively only perfectly hybridised and juxtaposed fragments are ligated .

In theory, the ligation possibilities between the oligonucleotides of the two modification libraries are threefold. Oligonucleotides from one or other of the modification libraries could ligate to oligonucleotides from the same modification library, or oligonucleotides from one modification library could ligate to those of the other modification library. If the added sequence has no alterations compared to the sequence used to produce the libraries then none of the oligonucleotides from the 3' end base replacement modification libraries should be able to ligate to each other since their 3' end bases have all been altered to bases other than those present at the corresponding position in the original sequence. Likewise, none of the oligonucleotides in the labelled 3' end base modification libraries should be able to ligate to each other because a free 3' hydroxyl is essential for ligation and this is absent on the dye labelled dideoxy terminators at this position. Similarly, ligation between oligonucleotides of the two different libraries should not be possible because the 5' position of the labelled 3' end base modification libraries will be juxtaposed to the mismatched 3' position of the end base replacement modification libraries.

Ligation should therefore only be possible if there is a single base difference between the original sequence used to make the modification libraries and the template sequence that is under investigation and exposed to the modification libraries under the hybridising conditions. Moreover, a ligation event will only occur when the end base replacement library contains a fragment whose end has been modified with a base change complementary to the base change present in the template. For example, if the sequence difference is C to T, then the end base replacement modification library which gave a change of G to A will contain an oligonucleotide which does not give a mismatch at the position of the sequence difference.

A ligation between two fragments of course results in a longer composite fragment. By subjecting hybridised and ligase treated modification library and template samples to a method of size analysis such as gel electrophoresis, any ligation products can readily be identified in terms of number of bases and further as necessary in terms of the nature of any label they carry.

The test described above is not dependent on how many oligonucleotides are present in a modification library, nor dependent on the length of the sequence of interest.

Figure 16 shows in general how template directed ligation will only occur between oligonucleotides when the oligonucleotides are annealed adjacently on the template and there are no mismatches between the ends to be joined and the template. Thus the 4 mer at position 5 cannot join to the 4 mer at position 4. Nor can the non-adjacent oligonucleotides ligate. This template is arbitrarily selected to be a non-mutant template. Comparison with Figure 17 shows how a mutation introduced into the original template now allows the 4 mer at position 5 to ligate to the 4 mer at position 4 because bases x and y are complementary.

Figure 18 corresponds to Figure 16 except that the general picture is substituted by real complementary bases. Similarly, Figure 19 corresponds to Figures 16 and 18 except that the instance of a mutant template is illustrated.

Consequence of a double stranded sequence

In the case of double stranded sequences of interest two modification libraries can yield ligation products since there will be a complementary change in the complementary sequence. In order to detect the nature of a base difference

between two otherwise identical sequences it is only necessary to examine the ligation products in mixtures which cover the complete range of end base replacement modification libraries. These could be with any one of the four labelled 3' end base modification libraries. Use in turn of all of the labelled 3' end base modification libraries allows the actual position of the sequence difference to be determined. This is possible because the 3' base at the end of each sequence word in a labelled 3' end base modification library is labelled according to its type. In this case each label is a different fluorescent dye. Examination of the dye present in the ligation products therefore identifies the base a fixed distance from the sequence difference and this distance is equal to the length of sequence words used in the particular labelled 3' end base modification library. Since the length of the sequence words differs between each labelled 3' end base modification library, each modification library identifies the base at a different distance from the sequence difference. Comparing the results obtained with each of the four labelled 3' end base modification libraries allows the actual four base sequence a fixed distance from the sequence difference to be read. Comparison to a known sequence will locate this actual position.

Use of just 48 modification libraries in liquid phase comprising oligonucleotides which have never been purified from each other, nor synthesised, but have been labelled in a manner which maintains referential integrity between the modification library subpopulations is enough to be able to determine both the nature of a sequence difference and to label its position in regard to sequence that has been determined to be a fixed difference from the sequence difference.

Use of double stranded sequences of interest allows sequences to be determined either side of a sequence difference because results are produced by the sense and anti-sense strands. The base change will never be the same

in each strand and so results will always be obtained with different end base replacement modification libraries. This has several advantages - for example it allows greater certainty in determining the position of a change and it allows changes to be confirmed by checking that the complementary change is observed in the opposite strand.

It is perfectly feasible to increase resolution further by increasing the number of different 3' labelled end base modification libraries that are used. Each modification library will differ according to the length of its sequence words so that using more lengths will result in more sequence being determined next to the site of the sequence change.

Sequence word length

The length of the sequence words used is a matter for one skilled in the art. In general, longer words are an advantage because they will have a greater fidelity in hybridisation. Longer words are less likely to be repeated in a sequence of interest and they are less likely to be altered to a sequence word which occurs elsewhere in the sequence of interest and thereby gives an ambiguous result. Similarly, the length of the sequence analysed will be determined by its yield of different sequence words.

In the modification library and detection systems described above ligation will only occur when sequence changes are present. This does not place a limit on the proportion of normal sequence to sequence variant in a sample under test, except insofar as it is possible to detect the ligation product.

The ligation method requires that the ligation products are detected and this is a matter for one skilled in the art. For example, capture of the end base replacement oligonucleotides can specifically capture the dye at the 3' ends of fragments in the 3' end base labelled library. Captured dye labelled fragments can be detected readily in gel systems for example. Similarly, separate fluorescent labelling of the end base replacement modification libraries would allow coincidence techniques to be used for detecting ligation.

Detection system

A general method has been described above with regard to the simplest sequence difference that can be envisioned. It is not in any way restricted. Multiple changes in a sample could be detected simply by increasing the number of modification libraries used, for example, by replacing bases in 2 positions per modification library would result in an increase in the level of discrimination provided by the library.

Detection of other types of sequence chance

The method described above is not restricted to the detection of base substitutions. It is perfectly feasible to design combinations of modification libraries that are able to detect deletions or insertions.

Deletions have the effect of bringing together regions of a sequence that otherwise are not adjacent. In the ligation test described above, deletion enables oligonucleotides that would not normally be juxtaposed to be able to ligate. The use of sequence word modification libraries each originating from a different start point at the end of the sequence of interest produces oligonucleotides that

cannot produce inter-modification library ligations when the modification libraries are mixed. This is because the ends of oligonucleotides from one modification library will never be juxtaposed to the ends from another modification library. Inter-modification library specific ligations are possible but can be prevented as described above by dideoxy terminators placed at their 3' ends. Provided that a deletion alters the reading frame of sample sequence it can be detected as a result of the ligation that it facilitates by bringing the sites of hybridisation of oligonucleotides from two different modification libraries into juxtaposition. Labelling of the oligonucleotides in the modification libraries as described above will allow the sequences at the limits of the deletion to be identified. This may bring about changes in length to the sequence words so it must be performed so that adventitious ligations are not facilitated.

Figure 20 shows an example of short arbitrary sequence and deletion modification libraries that can be produced from it. From top to bottom there is first a modification library produced originally by cyclical fragmentation 6 bases at a time from position 1 and then deletion each of the end bases, secondly there is a modification library produced originally by cyclical fragmentation 7 bases at a time from position 1 and then deletion of the 2 end bases from each fragment. The next two modification libraries correspond to the first two except that cyclical degradation commenced originally at position 2.

Figure 21 shows how an end base deleted fragment modification library can be used to detect a deletion. The site of a region to be deleted in a target is shown in the filled rectangle. Oligonucleotides from a deletion modification library are shown annealed to the target. The regions of their deletions are shown in broken line rectangles. They would not normally be able to ligate because they are not adjacent to each other. They are adjacent however in the deleted target shown

next with the shaded rectangle removed. Annealing to this target therefore allows them to ligate.

Figure 22 is a table of examples of possible additions and insertions in fragments from an arbitrary sequence. The first column shows the fragments. Next to each possible one base left hand end addition is shown. The third column shows each two base left hand end addition. Similarly, the fourth and fifth columns shows insertions of one and two bases respectively between positions 2 and 3.

Figure 23 illustrates how end base addition modification libraries can be used to detect an insertion. The arrow marks the position of a future insertion in the target. Fragments from an end base addition modification library are shown annealed to the target. The bases on their ends are shown displaced because they have no complementarity to the target and therefore cannot anneal. Next the target with an insertion is shown. Now the right hand oligonucleotide can completely anneal because the bases at its end are complementary to the insertion. It can therefore ligate to the right hand of the adjacent oligonucleotide shown next. Note the left hand end of the left hand oligonucleotide remains unannealed.

Insertions have the effect of separating the sites of hybridisation of oligonucleotides that would normally be in juxtaposition. Use of modification libraries that have end base additions will in the appropriate cases bridge the gap and allow ligation again. Measures must be taken to ensure that the effects of adventitious bridging between oligonucleotides that would normally be separated are taken into account.

<u>The construction of fragment modification libraries having referential integrity</u>

Figure 24 ((i) to (vii)) illustrates how information from fragment modification libraries can be combined to interrogate a multiplicity of samples having slight differences between their component polynucleotide templates. By way of example, arbitrary target sequences are listed first. The first sequence is chosen to be the "normal" reference sequence and the remainder have small changes. Sequence 2 for example has a G to A substitution while sequence 5 has a G deletion. The remainder of the table describes given oligonucleotide modification libraries derived from sequence 1 and the consequences on attempting to perform template directed ligation of the modification libraries in combination.

The first column lists 6 base sub-population fragments produced sequentially by moving 1 base further into sequence 1 each time. Column 2 summarises the fragments produced by producing all possible left hand end base substitutions. Column 3 shows the fragments produced by further processing the fragments in column 2 so that they are labelled according to the specific base at their right hand end in order 0, 1 and 2 bases further from the substituted bases. Column 4 (reading the sequences from top to bottom and not left to right) is the 4 base fragment library produced from sequence 1. Where an intersection occurs between column 3 and column 4 such that the libraries from 3 and 4 contain members that are able to ligate the intersection is marked with the number of the sequence template which allows the ligation. The oligonucleotides responsible can therefore be read off from the intersection. For example, the oligonucleotides GGAG from the 4 base library can ligate to the G to A substitutions from the GTATGG fragment of the substituted and labelled libraries. The sequence GTT commencing 5 bases from the substitution responsible can therefore be read from the labelled fragments ATATGG, ATATGGT and ATATGGTT. This serves to identify the nature and position of the change relative to sequence 1.

The result can be ambiguous. For example, the oligonucleotides ATGG from the 4 base library can ligate to the G to A substitutions from the GGTATGG fragment of the substituted and labelled libraries. The sequence GGT commencing 5 bases from the substitution responsible can therefore be read from the labelled fragments AGTATG, AGTATGG and AGTATGGT. This serves to identify the nature and position of the change relative to sequence 1 which superficially would appear to be a substitution. The change responsible however is a G deletion which brings oligonucleotides together that otherwise would be separated by a gap. It is necessary to examine other intersections for example the variants of GGTATG and their possible ligation with ATGG to obtain a complete picture.

Column 4 on Figures 24(iv) to (vii) also lists possible refinements of the 4mer libraries. In this case they have been divided into those starting at possible odd positions and those starting originally at possible even positions. These help to distinguish between possible alternatives that are responsible for a given ligation on the basis that the alternatives may have been partitioned between the libraries.

The following mathematical formulae describe in summary the constituent sub-populations which will be present in various types of modification library in accordance with the invention. Each modification library has the capability for detecting the nature and position of a particular change in nucleic acid sequence. Modification libraries can be combined with one another to provide for parallel interrogation of a multiplicity of polynucleotide sample species.

<u>1 base substitution modification library</u>

$$\sum_{n=1}^{\infty} S_{n,c}$$

where S_n is a strand of L consecutive bases starting at position n, where L could be the minimum to maximum possible length generated from the target of interest and c is the position of a single base substitution.

2 base substitution modification library

$$\sum_{n=1}^{\infty} \sum_{c=1}^{L} \sum_{f=1}^{\infty} S_{n,c,f}$$

where Sn is a strand of L consecutive bases starting at position n, where L could be the minimum to maximum possible length generated from target of interest, c is the position of a substitution and f is the number of bases substituted from position c.

Insertion modification library

$$\sum_{n=1}^{\infty} \sum_{j=1}^{L} \sum S_{n,j}$$

where Sn is a strand of L consecutive bases starting at position n, where L could be the minimum to maximum possible length generated from the target of interest, j is the position of an insertion of a specified number of bases.

Inversion modification library

$$\sum_{n=1}^{\infty}\sum_{b=1}^{L}\sum_{w=1}^{L}S_{n,b,w}$$

where Sn is a strand of L consecutive bases starting at position n, where L could be the minimum to maximum possible length generated from the target of interest b is the first of w number of bases the sequence of which is inverted with regard to the original sequence at that position within the fragment

Deletion modification library

$$\sum_{n=1}^{\infty}\sum_{d=1}^{L}\sum S_{n,d,f}$$

where Sn is a strand of L consecutive bases starting at position n, where L could be the minimum to maximum possible length generated from the target of interest, d is the position of a deletion of a specified number of bases where f is the number of bases deleted consecutively and including position d.

In order to obtain formulae describing combinations of libraries simply add the required series together:

$$\sum S_{n}+\sum_{n=1}^{\infty}\sum_{c=1}^{L}S_{n,c\ldots}$$

(This formula represents adding final modification libraries together. It is not the same as doing the molecular biology in combination because regarding the latter order is important where as with the former it is not. A modification library produced by a substitution followed by an addition is not necessarily the same

as one produced by an addition followed by a substitution. Mixing two or more modification libraries always has the same effect regardless of the order of mixing).

Example 1 - Determination of the identity and position of a change in a given sequence using liquid libraries and selected model sequences as samples: I

These examples concern the arbitrary sequence

1. 5' gggatcttgtcgaataaagtcgaggtgctagtttcataagcaaa (SEQ ID NO: 4)

and its variations :

2. 5'gggatcttgtcgaataaagtctaggtgctagtttcataagcaaa (SEQ ID NO: 5), a G to T substitution,

3. 5'gggatcttgtcgaataaagtccaggtgctagtttcataagcaaa (SEQ ID NO: 6), a G to C substitution,

4. 5'gggatcttgtcgaataaagtcaggtgctagtttcataagcaaa (SEQ ID NO: 7), a G deletion,

5. 5'gggatcttgtcgaataaagtcgaaggtgctagtttcataagcaaa (SEQ ID NO: 8), an A insertion and

6. 5'gggatcttgtcgaataaagtcgatggtgctagtttcataagcaaa (SEQ ID NO: 9), a T insertion.

They illustrate principles involved in using liquid libraries of oligonucleotides to determine some feature of a sequence. In this case sequence variations are being determined. Oligonucleotide fragments of the sequence were produced to determine the positions and nature of the variations. All oligonucleotides were supplied by Oswell (Southampton). They were synthesized chemically and purified by HPLC. Since advantage could be taken of chemical synthesis for the relatively small number of oligonucleotides required it was preferred to label at the 5' end according to the specific base found there. Similarly, the varied bases

were produced on the 5' end of the unlabelled oligonucleotides. This example does not attempt to describe the production and use of comprehensive liquid libraries which are illustrated in examples which follow.

The basis of the assay was to determine whether template dependent ligation of test oligonucleotides had occurred according to sequence variations found in the template and whether this was the only significant reaction when other near identical competing reactions which would have allowed mismatching at the point of ligation were possible . Oligonucleotides were therefore varied in size so that they could be distinguished following ligation by analysing the products through gel electrophoresis. Oligonucleotides were labelled with fluorescent dyes so that they could be detected using a fluorescent sequencer. Unlabelled oligonucleotides had a phosphate at their 5' ends to provide the means for ligation. Non of the 3' ends of any of the oligonucleotides were blocked to prevent ligation since the sites of hybridisation under investigation should only have permitted ligation between an oligonucleotide of the labelled set and an oligonucleotide of the 5' phosphorylated set. The instrument used for analysis in the examples described is the Model 377 supplied by Perkin Elmer operated according to the instructions of the manufacturer. This allows four different dyes to be distinguished in one sample. Fragments produced were sized using the gene scan software and the size standards of the manufacturer. A convention of FAM to label 5' terminal A, JOE to label 5' terminal C, ROX to label 5' terminal G and TAMRA to label 5' terminal T was adopted.

The probe oligonucleotides used were :

Oligo ID	Sequence	SEQ ID NO:
A	5'agactttattcga	10

B	5'cgactttattcgac	11
C	5'ggactttattcgaca	12
D	5'tgactttattcgaca	13
E	5'agactttattcgac	14
F	5'cgactttattcga	15
G	5'ggactttattcgac	16
H	5'tgactttattcgac	17
I	5'agactttattcgaca	18
J	5'aactttattcgacaag	19
K	5'cactttattcgacaag	20
L	5'gactttattcgacaag	21
M	5'tactttattcgacaag	22
N	5'acgactttattcgacaag	23
O	5'ccgactttattcgacaaga	24
P	5'gcgactttattcgacaagat	25
Q	5'tcgactttattcgacaagatc	26
R	5'atcgactttattc	27
S	5'ctcgactttattcga	28
T	5'gtcgactttattcga	29
RS-U	5'ttcgactttattcga	30

Oligos A to T and RS-U were all 5' phosphate.

Oligo ID	Label	Sequence	SEQ ID NO:
U	5'TAMRA	5'tatgaaactagcacct	31
V	5'FAM	5'atgaaactagcacct	32
W	5'TAMRA	5'tgaaactagcacct	33
X	5'ROX	5'gcttatgaaactagcacc	34

Y	5'JOE	5'cttatgaaactagcacc	35
Z	5'TAMRA	5'ttatgaaactagcacc	36

Different combinations of oligonucleotides were mixed with the template or a variation of the template, heated to 72 °C to denature secondary structures and then allowed to cool to room temperature to anneal. The dye labelled oligonucleotides were pooled in equimolar amounts and used together. Reactions were performed in 20 microlitre volumes containing 0.2 units of T4 DNA ligase, x1 ligase buffer, 120 pmoles of template, 24 pmoles of the labelled oligonucleotides (4 pmoles each) and 24 pmoles of the unlabelled oligonucleotide mixtures (3 pmoles each). Ligase buffer and T4 DNA ligase were both supplied by Boehringer Mannheim. Equivalent units of T4 DNA polymerase from NEW England Biolabs were substituted on occasion without apparent change to the results. 30 minutes were allowed to elapse following addition of the ligase buffer before ligase was added to allow oligonucleotides to anneal before they had an opportunity to ligate to each other. Ligation reactions were performed for 16 hours at 37 °C. Products of the ligation reactions were precipitated by 2.5 volumes of ethanol and 0.1 volume of 3M sodium acetate pH5.2. Precipitates were collected at 13,000 x gravity for 30 minutes. Ethanol was aspirated away and the pellets were washed by vortexing with 70 % ethanol. Re-pelleting was achieved at 13,000 x gravity for 5 minutes, the ethanol aspirated and the pellets dried at 37 °C for 15 minutes. Pellets were dissolved in 3 microlitres of sequencing gel loading buffer comprising deionised formamide containing 5mM ethylene diamine acetic acid at pH 8 and 10 mg/ml Blue Dextran 2000. 1.5 microlitres of the dissolved samples were analysed by polyacrylamide gel electrophoresis using the 377 instrument described above. Samples were diluted as necessary in sequencing loading buffer prior to gel loading to achieve the required signal strengths. The intention was to achieve a 10 to 100 fold

excess of template over any given oligonucleotide to which it would be hybridized to avoid crowding of oligonucleotides on the templates. In practice, lower ratios could be used satisfactorily. Typically, 12-120 pmole of template or its variation were used in a 20 microlitre ligation reaction at a 30-40 fold excess over any given oligonucleotide. These considerations were thought to be important should more complex mixtures be used since too much template could result in oligonucleotides being juxtaposed too infrequently for sufficient ligation to occur. Alternatively, crowding could occur resulting in steric exclusion from templates. Larger reactions were used if necessary to achieve the desired signal strength although in general detection was easy and concentration by precipitation was not necessary. Oligonucleotides that were to be examined for their ability to ligate to each other as directed by the template were used in equimolar amounts. Different ratios of template or its variation were used to maximize the efficiency with which oligonucleotides were hybridized in juxtaposed positions on the template without crowding overlapping sites.

Maximum signal strengths were obtained as expected with the input template and the oligonucleotides used. Using the original sequence 1. with the labelled oligonucleotides and unlabelled oligonucleotides A, B, C, D and J, K, L and M respectively gave TAMRA, FAM and TAMRA products of 30, 29 and 28 bases respectively. This corresponds to an end sequence of tat the length of the corresponding original labelled oligonucleotides away from the point of ligation and points to oligonucleotide B as the unlabelled oligonucleotide involved since this has the 14 bases required to make up the difference. It also suggests that there were no significant ligations to any of the oligonucleotides J,K, L and M since TAMRA and FAM products of 32 and 31 bases respectively would have resulted. W can give a 30 base TAMRA product as observed but this is not consistent with the position of the FAM product. These results are obtained

despite the displacement of the 3' ends of J to M by only one base relative to that of B at the site of ligation indicating that oligonucleotides need to be exactly juxtaposed for template directed ligation to occur. They also rule out significant ligation by the oligonucleotides A, C and D which entirely correspond to the ligation site but have unpaired (mismatched) bases at their 5' ends. c at the 5' end of B is therefore suggested as the base at the point of ligation.

Use of oligonucleotides E,F, G, H or F,G,H and I in place of A to D gives similar results and rules out any size biases affecting the outcome of ligation. The TAMRA, FAM and TAMRA products of 29, 28 and 27 bases are consistent with ligation of the labelled oligonucleotides occurring to F, again pointing to c at the point of ligation. The tat found within the tttattc sequence is also a candidate site for ligation but in this case the c base required at the point of ligation would be off the end of the target.

Labelled oligonucleotides X and Y fail to produce ROX and JOE labelling following ligation as expected from the position of their 5' ends one base removed from the point of ligation.

Use of the variant sequence 2. as the template produces products consistent with the participation of A, E and I as the unlabelled oligonucleotides in ligation i.e. TAMRA, FAM and TAMRA products of 29, 28 and 27 bases respectively were produced with A, TAMRA, FAM and TAMRA products of 30, 29 and 28 bases respectively were produced with E and TAMRA, FAM and TAMRA products of 31, 30 and 29 bases respectively were produced with I. This identifies a as the new point of ligation and a difference of g to t between the two template strands.

Note that prior knowledge of the two template strands is not necessary to determine the relative positions and nature of the base differences between the two templates. It is sufficient to know that the probe oligonucleotides are derived from the templates, their length, their relative positions to each other and the nature of the differences at their altered ends and the actual base at the opposite end to their changes. In the absence of any sequence information about the templates it could still be determined that a base change had occurred at the point marked by the ends of the oligonucleotides, the nature of the change and the relative position of the change.

Use of the variant sequence 3. as the template produces products consistent with the participation of C and D and G and H as the possible unlabelled oligonucleotides in ligation. This identifies g or t as the new point of ligation and a difference of g to a or c between the template strands. C and D can be distinguished when used in isolation but the point of liquid libraries is to distinguish between competing possibilities when used together in combination.

Note that the oligonucleotides require a template to which their ends at the point of ligation are complementary.

Use of the variant sequence 4. with A to D and gives ROX, JOE and TAMRA products of 33, 32 and 31 bases respectively, E to H gives ROX, JOE and TAMRA products of 32, 31 and 30 bases respectively and J to M gives TAMRA, FAM and TAMRA products of 32, 31 and 30 bases. This is consistent with one of J to M participating in ligation and a deletion of 1 base (g) having occurred at the point of ligation.

Use of the variant sequence 5. with A to D and N to Q gives TAMRA, FAM and TAMRA products of 37, 36 and 35 bases which is consistent with Q participating in ligation and an insertion of 1 base (a) having occurred at the point of ligation. Significantly, Q bridges the extra base to U, V and W.

Use of the variant sequence 6. with A to Q does not produce any significant products. U to W are juxtaposed to N to Q but U to W are not paired correctly at their ends with template 6. The set R, S, T, RS-U restores ligation producing 1 base shorter ROX, JOE and TAMRA products, respectively. This is consistent with X, Y and Z providing the labels on ligation. The 27 to 29 base fragments indicate an a bridging the gap of a t insertion on template 6 compared to template 1 at the point of ligation.

ROX, JOE and TAMRA products of 39, 38 and 37 bases respectively are produced with set N to Q and template 1 as expected if Q ligates to X, Y and Z.

Example 2 - Construction of liquid libraries corresponding to exon 5 of human p53

Fragments containing human p53 exon 5 were produced by PCR amplification from human placental genomic DNA supplied by Sigma, using the primers 5'ttccagttgctttatctgttca (SEQ ID NO: 37) (position 12988-13020 on the genomic sequence) and 5'aagagcaatcagtgaggaatcaga (SEQ ID NO: 38) (position 13293-13317 on the genomic sequence). The amplified products were TA cloned into the plasmid pCR2.1 (In Vitrogen). 500ml cultures of *E.coli* containing the correct constructs (identified by dRhodamine dye terminator sequencing - Perkin Elmer) were used to produce plasmid DNA (Qiagen). Fragments were

produced from the p53 gene within the plasmid by a cyclical process of ligation and cutting. An adaptor containing the site for a Type11s restriction endonuclease in this case BpmI was added to the end of p53 containing fragments during the ligation step. The position of the restriction site ensured that during the cutting step a fixed number of bases was added to the adaptor from the end of the p53 fragment. This exposed the end of the fragment for further cycles of ligation and cutting. Plasmid was pre digested in the vector with BpmI or PvuI as appropriate before cyclical cutting and ligation to determine the start points of the process. Typically, 7.5 micrograms of plasmid DNA were digested with 20 units and 25 units of BpmI and PvuI (both New England Biolabs) respectively in 200 microlitres reactions comprising x1 New England buffer 3. Reactions were performed at 37 °C for 2 hours, the enzyme additions were repeated and incubation continued for a further 2 hours. This relatively harsh treatment proved necessary with these enzymes and this plasmid but reflects the starting materials rather than the process itself. DNA was purified by extracting twice with 1:1 phenol / chloroform and then passing through an S-200 Microspin column supplied by Pharmacia and used according to the manufacturers instructions.

Prior to cyclical cutting and ligation, adapters containing an appropriately positioned Bpm1 site were used to determine starts points for the process beyond the original ends produced by the original Pvu1 and Bpm1 cuts. A second Bpm1 cut could then remove a predetermined number of bases to produce a new start point. The adaptor was double stranded with a 2 base 3' overhang chosen to recognise the end from which cleavage should occur. At least 30 pmoles of adaptor are ligated to 1 pmole of construct in reactions containing not more than 200 pmoles total per 50 microlitres. 1 unit of ligase (Boehringer) was used according to the manufacturers conditions. Intermediate

purification is by silica columns (Qiagen) used according to the manufacturers instructions. 2 units of Bpm1 were used per 20 microlitres of reaction for 2 hours at 37 °C twice.

For the cyclical cutting and ligation, 0.25 microgram amounts of pre digested and purified DNA were ligated and cut in 25 microlitres of x1 ligase buffer supplied by Boehringer. 0.5 units of ligase (Boehringer) are added at the start of the reaction. Incubation was initially at 14 °C for 40 minutes. 3.75, 15 or 60 pmoles of adaptor were used per 25 microlitres of reaction. Adapters were of the design: 5'ccagtcgcaggtctcaagctcgacagctggag(v)nn (SEQ ID NO: 39) with the corresponding antisense strand commencing 2 bases in from the 3' end to leave a 2 base 3' overhang on the sense strand described. V is a variable number of predetermined nucleotides between 0 and 14 and is chosen to achieve removal of the desired number of nucleotides from the construct on digestion with the BpmI Type11s restriction endonuclease. Adapters were synthesised chemically and purified by HPLC. All adapters were prepared and supplied by Oswel (Southampton). Adapters for each size class of interest were synthesised as four syntheses to the general design of na, nc, ng or nt rather than nn and then mixed in equimolar amounts to achieve nn to minimise biases introduced by differential incorporation rates of mixed synthesis. Prior to use, adapters were mixed with an equimolar amount of their antisense strand, heated to 72 °C for 10 minutes to reduce secondary structure and then allowed to anneal at room temperature.

40 minutes after the start of ligation, 1 unit of BpmI (New England biolabs) was added to the ligations and the reaction continued at 37 °C for 40 minutes. Temperature cycles of 40 minutes at 14 °C followed by 40 minutes at 37 °C were continued throughout the reactions. The activity of the BpmI reduces

significantly during the reaction so fresh enzyme (1 unit) was added at 1 hour intervals up to 6 hours. Cycling was continued for 16 hours.

Samples were taken every 2 hours up to 6 hours and after 16 hours and analysed by agarose and also by denaturing polyacrylamide gel electrophoresis to monitor the extent of reaction. Gels were stained using Vistra Gold at 1 microgram/ml (Amersham) and visualised using a fluorimager (Molecular Dynamics). Agarose gels contained 3% Nusieve 3:1 (FMC) and tris acetate electrophoresis buffer. Polyacrylamide gels contained 15 % polyacrylamide, 6M urea and TBE electrophoresis buffer. Gel running conditions were contemporary and as described by the suppliers (FMC) or in Maniatis. Controls included omission of ATP, omission of ligase, a zero time point, omission of adapters and omission of the restriction endonuclease. Greatest yields of the desired products were obtained with the longest incubation times and the highest concentrations of adapters. No products were obtained unless all of the reaction components were present. Omission of the adapters allowed autoligation of the plasmid, demonstrating that the adapters were ligating to the plasmid ends. The size of the products was entirely consistent with the design of the input adapters. This result is extremely significant since it is not obvious that cyclical cutting and ligation would be possible as they are opposing processes with differing requirements. Double digests of the plasmid with PvuI and BpmI showed that the process could reduce the sizes of the resultant fragments by at least hundreds of bases making it amenable to targets of significant length. Cyclical cutting and ligation at high efficiency is also valuable to make maximum use of the substrate and also to ensure libraries with optimum representation. Type11s restriction endonucleases have been reported to lack fidelity in terms of the site at which they actually cut as opposed to where they are expected to cut. This was not significantly our experience with BpmI. Fragments of interest could be

purified by standard procedures following separation by denaturing polyacrylamide gel electrophoresis. Labelling on their 3' end could be achieved by ligating on a second adaptor containing a Type11s site for FokI which left a 4 base 5' overhang at the end to be labelled. Base specific labelling was achieved through the dideoxy terminator core cycle sequencing kit (Perkin Elmer). The reaction was not cycle, the unlabelled nucleotides were omitted and the labelled nucleotides were titrated to identify the amounts which gave most even use of all four labelled bases. Removal of unwanted adaptor fragments could be achieved by including a second site for BpmI in the first adapters. This site was situated close to the end of the adapters so that when used it would exactly cut the adaptor from fragments of interest. The site was inactive during cyclical cutting and ligation because in our hands, 26 bases were needed in front of the 5' end of the BpmI site before efficient cutting could occur. These were not made available in the adapters used so that use of the site could be made active when required. The adapters had a 4 base 5' overhang which allowed a second adaptor to be ligated on and supply the necessary bases to allow the second BpmI site to be used. Appropriate positioning of the second BpmI site also allowed selected bases at the end of the fragments of interest to be exposed so that they could be identified by ligation to a second adaptor. This second adaptor also contained a BpmI site so that it could be removed together with bases to be replaced at the 3' end of the fragments of interest. When the second adaptor was used it was double stranded. It had the general design w26Type11synn1x. w26 was a predetermined sequence of at least 26 bases, Type11s was the site for the Type11s restriction endonuclease (in this case Bpm1) yn was a predetermined sequence which ensures that the restriction endonuclease cut in the desired place to remove the selected bases, n1 was any one of all four possible bases at the penultimate 3' position in any given population of the adapters and x was one of the 4 bases a, c, g or t chosen to

ensure ligation to the fragments with the corresponding ends of choice in the library of fragments. n1x formed a 2 base 3' overhang. The second adapters were used together with a population of identical adapters except that x was replaced by each of the other 3 unused bases so that all other possible 3' ends could also be recognised by ligation. A base alteration was also made in the Type11s site so that the restriction endonuclease was no longer active on these adapters. The purpose of these adapters was to ensure that ends of fragments not having the base of interest would be blocked by ligation and take no further part in the process. Following cleavage of the second adapters a third set of adapters were ligated to the resultant fragments. These were of the same general design as the second set of adapters except that their 3' ends were of the form xn1n1 where, x was double stranded. n1n1 was single stranded and corresponded to a base or combination of bases (other than the one already removed) that was to be added back to the 3' end of the fragments of interest. yn was chosen to ensure that following cleavage with the Type11s restriction endonuclease the desired base(s) were added back to the fragments of interest.

Fragments were purified at intermediate stages of the process by separation through and then extraction from denaturing polyacrylamide gels as above. Ligations were performed at a molar ratio of 30 adapters to 1 of fragments with 1 unit of ligase (Boehringer) per 50 microlitres of reaction with up to 200 pmoles total per 50 microlitres. Incubation was at 14 °C for 16 hours. BpmI was used at 2 units per 20 microlitre reaction for 2 hours at 37 °C twice per reaction. Extent of reactions were monitored by gel electrophoresis as above and repeated as required.

Fragment libraries prepared in these ways in these ways were purified following polyacrylamide gel electrophoresis and used in conjunction with the original

fragments to identify substitutions in the p53 gene of exon 5 using the approaches described in the first example.

Example 3

It was desired to create a system that would be universally applicable. Four series of vectors were created for this purpose. The first series pFRAGnn allowed any region of interest to be cloned such that the start point for producing fragments from the cloned region could be any point between the third and sixteenth base from a given end of an insert. The vector used determines the particular positions. pFRAG03 for example causes fragments to be produced from the third base. Each successive vector moves this point one base into the insert until pFRAG16 which cuts at the sixteenth base.

We had found that the efficiency of the cutting and ligation process was reduced as it progressed away from the original start point. The representation of the fragments produced was therefore not uniform. This was addressed in the design of the pFRAGnn vectors. There are two Bpm1 in each vector. The sites are placed so that they face eachother from opposite sides of the insert. Bpm1 can then be used to excise the insert. Cyclical cutting and ligation can then occur from both ends of the insert to even out representation of the fragments produced. The Bpm1 sites are also placed in each vector so that as cyclical cutting and ligation progresses it cuts at the same positions regardless of the direction from which it reached that position ie the same fragments are produced from either side. It is therefore necessary to use inserts whose length also allows this to occur.

Directional cloning is achieved by having two sites in the vector for the type11s restriction endonuclease Bbs1 where each site produces a different cohesive end. The fragment containing the Bbs1 recognition sequence is removed from the vector on cloning. Bbs1 produces the same types of cohesive ends with each vector in the series that all of the vectors can clone the same fragments. PCR was used to add the corresponding sets of 4 bases to the ends of the inserts. This creates a universal system since any PCR primers can easily be designed to have extra bases at their 5 prime ends so that they are incorporated into the final product. The bases are rendered single stranded after PCR by the combined action of the 3' exonuclease and DNA polymerase activities of T4 DNA polymerase. In practice 5 bases are added to the 5 prime ends of the PCR primers and one type of base was excluded from the first four positions but included at the fifth position. The action of T4 DNA polymerase plus the single deoxytrinucleotide corresponding to the fifth position then sets up a futile cycle whereby the exonuclease removes all five 3' bases but is able to add back the fifth base. This base is repeatedly added and removed with the net effect that the 5 prime single stranded end required for cloning is produced.

We produced fourteen different versions of the pFRAGnn vectors. Full use of the distance between the recognition site and cutting sites of Bpm1 could then be used to expose all possible positions up to 14 bases for cyclical cutting and ligation.

The remaining three series of vectors concern manipulation of the fragments produced by cyclical cutting and ligation. In general the vectors have Bpm1 sites adjacent to a sequence for capturing by ligation any given double stranded fragment having any particular 3' dinucleotide single stranded ends. Each series

of vectors has the Bpm1 recognition site a different distance from the point of capture to suit the manipulations that followed capture.

The pSELECTnn vectors allow fragments to be captured according to the actual single stranded, dinucleotide sequence at their 3' end. Their Bpm1 site is positioned to exactly release the captured fragments. The pRESECTnn vectors are similar to the pSELECTnn vectors except that the Bpm1 site serves to remove the most 3' captured nucleotide on release of the captured fragment. pREPLACEXnn vectors have a Bpm1 site situated to add a base back to the 3' end of captured fragments when they are released by Bpm1.

The parent plasmid for all of the vectors was pUC19. This has a unique site for Bpm1 and two sites for the type11 restriction endonuclease BsrD1 all in its ampicillin gene. The Bpm1 site would have interfered with our process and it was useful also to be able to use BsrD1. We therefore used a laboratory vector pIND10 from which the Bpm1 and BsrD1 sites had been removed from the ampicillin gene by recombination PCR using primers that substituted bases in the recognition sites for these enzymes. The substitutions were neutral so that the ampicillin gene retained its ability to confer resistance to the antibiotic.

It had been intended to produce this plasmid by a tri molecular recombination. A reverse primer 5' TCTCAACAGCGGTAAGATCC (SEQ ID NO: 59) or 5' CAACAGCGGTAAGATCCTTG (SEQ ID NO: 60) beyond the Xmn1 site and a forward primer 5' ACGCTCACCGGCACCAGATT (SEQ ID NO: 61) or 5' TCACCGGCACCAGATTTATC (SEQ ID NO: 62) with a mismatch to the Bpm1 site were used to PCR one region. A reverse primer 5' CCTGTAGCTATGGCAACAAC (SEQ ID NO: 63) or ATGCCTGTAGCTATGGCAAC (SEQ ID NO: 64) or 5'

TGCCTGTAGCTATGGCAACA (SEQ ID NO: 65) with a mismatch to the BsrD1 site at 1926 bases on pUC19 and a forward primer IXBP 5' AGTATTTGGTATCTGCGCTC (SEQ ID NO: 66) or 5' TATTTGGTATCTGCGCTCTG (SEQ ID NO: 67) or 5' TTGGTATCTGCGCTCTGCTG (SEQ ID NO: 68) or 5' GTATCTGCGCTCTGCTGAAG (SEQ ID NO: 69) at 1306 bases PCRd a second region. A third region was amplified using a forward primer 5' GGTAATACGGTTATCCACAG (SEQ ID NO: 70) or 5' CAACAGCGGTAAGATCCTTG (SEQ ID NO: 71) spanning the Afl11 site and a reverse primer 5' CTCGCGGTATAATTGCAGCA (SEQ ID NO: 72) or 5' TCTCGCGGTATAATTGCAGC (SEQ ID NO: 73) or 5' GTCTCGCGGTATAATTGCAG (SEQ ID NO: 74) with a mismatch to the BsrD1 site at 1748 bases. PCRs corresponding to all possible combinations of alternative primers were used and the best products used to continue the process. Reactions were performed in all cases with 0.1 to 1 ng of pUC19 PCR at 94.5°C for 5 minutes, then 32 cycles of 94.5°C and 65°C for 30 seconds each and 72°C for 1 minute with a 5°C gradient at the 65°C step. A final incubation of 72°C for 10 minutes was performed. 50ul reactions containing 0.2mM dNTPs, 25 pmoles of each primer, 2.5 units of AmpliTaq Gold (Perkin Elmer) 2.5mM $MgCl_2$.

PCR products from the three regions were purified through a QIAquick spin column (Qiagen), serially diluted 1 to 2 each in water and the dilutions of all three products mixed in equal proportions in all combinations. PCR was repeated using the Xmn1 reverse and the Afl111 forward primers according to the conditions above. It was anticipated that the desired region would only be able to amplify if the three regions were initially extended using the corresponding parts of the other regions as templates. Once a region spanning the two primer

sites had been achieved then amplification of the whole region would ensue combining the mutations originally incorporated during amplification of the first three regions. The PCR products were purified as above, cut to completion with Xmn1 and Afl111 purified as above and ligated to similarly cut and purified pUC19 (all New England Biolabs). A first set of 16 vectors was produced. PCR product was used at a 3 molar excess to pUC19 and 0.2 ug of pUC19 were used per 20 ul ligation containing 0.2 units of T4 DNA ligase. Ligations were used to transform E.coli XL1-Blue by the $CaCl_2$ method of Hannahan. Ampicillin at 50 ug / ul was used as counter selection. Plasmids were prepared using QIAprep 96 and analysed by cutting to completion with Bpm1 and BsrD1 and analysing by agarose gel electrophoresis. Double digests with Xmn1 were performed to confirm the positions of any differences compared to pUC19. In practice only the Bpm1 mutation was incorporated. The experiment was therefore repeated except that only 2 regions were amplified initially and the new plasmid which had lost its Bpm1 site was used as the target. The first region used the Xmn1 reverse primer and a BsrD1 forward primer 5' TGCTGCAATTATACCGCGAG (SEQ ID NO: 75) or 5' CTGCAATTATACCGCGAGAC (SEQ ID NO: 76) which incorporated a substitution into the BsrD1 site at 1748 bases. The second region used the Afl111 reverse primer and the reverse primer which incorporated a substitution into the BsrD1 site at 1926 bases. The two regions were combined and transformed as before. Screening as before yielded several plasmids which had lost all three sites. PIND10, one of these plasmids was used for further work.

The polylinker of the vector pIND10 was removed by cutting 10 ug to completion with EcoR1 and HindIII (New England Biolabs) in a 100ul reaction and purification through a QIAspin column (Qiagen). The polylinker was replaced by oligonucleotides EX3_01 to 16 (SEQ ID NOs: 77-92 respectively):

EX3_01 5' aattctggagaacattgccgacaaggatcc
EX3_02 5' aattctggagaccattgccgacaaggatcc
EX3_03 5' aattctggagagcattgccgacaaggatcc
EX3_04 5' aattctggagatcattgccgacaaggatcc
EX3_05 5' aattctggagcacattgccgacaaggatcc
EX3_06 5' aattctggagcccattgccgacaaggatcc
EX3_07 5' aattctggagcgcattgccgacaaggatcc
EX3_08 5' aattctggagctcattgccgacaaggatcc
EX3_09 5' aattctggaggacattgccgacaaggatcc
EX3_10 5' aattctggaggccattgccgacaaggatcc
EX3_11 5' aattctggagggcattgccgacaaggatcc
EX3_12 5' aattctggaggtcattgccgacaaggatcc
EX3_13 5' aattctggagtacattgccgacaaggatcc
EX3_14 5' aattctggagtccattgccgacaaggatcc
EX3_15 5' aattctggagtgcattgccgacaaggatcc
EX3_16 5' aattctggagttcattgccgacaaggatcc

and their complementary strands respectively EX3_17 to 32 (SEQ ID NOs: 93-108) having the general formula of SEQ ID NO: 109 (5' agctggatcc ttgtcggcaa tgnnctccag):

EX3_17 5' agctggatccttgtcggcaatgttctccag
EX3_18 5' agctggatccttgtcggcaatggtctccag
EX3_19 5' agctggatccttgtcggcaatgctctccag
EX3_20 5' agctggatccttgtcggcaatgatctccag
EX3_21 5' agctggatccttgtcggcaatgtgctccag
EX3_22 5' agctggatccttgtcggcaatgggctccag

EX3_23 5' agctggatccttgtcggcaatgcgctccag

EX3_24 5' agctggatccttgtcggcaatgagctccag

EX3_25 5' agctggatccttgtcggcaatgtcctccag

EX3_26 5' agctggatccttgtcggcaatggcctccag

EX3_27 5' agctggatccttgtcggcaatgccctccag

EX3_28 5' agctggatccttgtcggcaatgacctccag

EX3_29 5' agctggatccttgtcggcaatgtactccag

EX3_30 5' agctggatccttgtcggcaatggactccag

EX3_31 5' agctggatccttgtcggcaatgcactccag

EX3_32 5' agctggatccttgtcggcaatgaactccag

Equimolar amounts of each complementary pair were mixed and heated to 95°C before cooling to ambient. Paired oligonucleotides were ligated for 16 hours at 16°C at a 3 molar excess to 200ng of the cut vector in a 20ul reaction containing 0.2 units of T4 DNA ligase (Boehringer). Ligated material was cut with 20 units of the restriction endonuclease Xba1 for 1 hour at 37°C to select against the original plasmid. E.coli XL1-Blue were transformed with the Xba1 treated material by the $CaCl_2$ method (Hannahan) and selected using 50 ug per ml of ampicillin and IPTG / X-gal blue white colour selection (0.004 % weight : volume each. Dark blue and pale blue colonies (3 each per transformation) were picked and screened for the new polylinker. Plasmids were produced from the colonies using QIAprep 96 (Qiagen) and cut with HindIII, BsrD1 and EcoR1 to score for the new polylinker. Candidate plasmids were diluted 1 in 1000 of water and amplified by PCR using the primers 5' AGGCACCCCAGGCTTTAC (SEQ ID NO: 110) and 5' CCGCACAGATGCGTAAGG (SEQ ID NO: 111) PCR products were purified by QUIAquick 96 and their sequence confirmed by dRhodamine dye terminator sequencing on the ABI 377 (Perkin Elmer) using the PCR primers as sequencing primers. Confirmed plasmids were prepared in bulk using the

QIAfilter plasmid maxiprep kit (Qiagen). These plasmids were numbered pINDnn. The vectors in all of our series are conventionally represented with the EcoR1 site of the original pUC19 plasmid on the left and the Hind111 site on the right. nn corresponded to the dinucleotide immediately adjacent to their Bpm1 site in the direction of the BsrD1 site so that the 2 base 3' overhang produced by BsrD1 corresponded exactly to the two particular bases found at nn. For example pINDag has the dinucleotide ag in its upper 3' single stranded end produced on digestion by BsrD1. The vectors in all of our series are conventionally represented with the EcoR1 site of the original pUC19 plasmid on the left and the Hind111 site on the right. In general plasmids that produced pale blue colonies after 24 hours at 37°C had the required polylinker sequences.

The plasmids pINDnn were used to produce the four further series of vectors. For the next three series the region between BsrD1 and EcoR1 of the pINDnn plasmids was replaced by cutting to completion with the restriction endonucleases EcoR1 and BsrD1. The methods used were the same as those described above except for the design of the oligonucleotides inserted.

The series of vectors pSELECTnn were produced by the oligonucleotides EX3_33 to EX3_48 (SEQ ID NOs: 112-127) replacing the EcoR1 to BsrD1 region of pINDaa to pINDtt respectively :

EX3_33 5'aattcctggag(n_{14})aa
EX3_34 5'aattcctggag(n_{14})ac
EX3_35 5'aattcctggag(n_{14})ag
EX3_36 5'aattcctggag(n_{14})at
EX3_37 5'aattcctggag(n_{14})ca
EX3_38 5'aattcctggag(n_{14})cc

EX3_39 5'aattcctggag(n_{14})cg

EX3_40 5'aattcctggag(n_{14})ct

EX3_41 5'aattcctggag(n_{14})ga

EX3_42 5'aattcctggag(n_{14})gc

EX3_43 5'aattcctggag(n_{14})gg

EX3_44 5'aattcctggag(n_{14})gt

EX3_45 5'aattcctggag(n_{14})ta

EX3_46 5'aattcctggag(n_{14})tc

EX3_47 5'aattcctggag(n_{14})tg

EX3_48 5'aattcctggag(n_{14})tt

and their complementary strand (SEQ ID NO: 128):

EX3_49 : 5' (n_{14})ctccagg

The reading frame of the lacZ alpha fragment of the vector was maintained. The string of 14 n's is to ensure that the Bpm1 site cuts the capture dinucleotides to exactly release any captured fragment. All 16 possible vectors were produced. The complementary strand produced a 5' aatt overhang and the appropriate 3' overhang for insertion into the particular pINDnn vectors.

The series of vectors pRESECTnn were produced by the oligonucleotides EX3_50 to EX3_65 replacing the EcoR1 to BsrD1 region of pINDaa to pINDtt respectively (SEQ ID NOs: 129-144):

EX3_50 5'aattccctggag(n_{13})aa

EX3_51 5'aattccctggag(n_{13})ac

EX3_52 5'aattccctggag(n_{13})ag

EX3_53 5'aattccctggag(n_{13})at

EX3_54 5'aattccctggag(n_{13})ca

EX3_55 5'aattccctggag(n_{13})cc

EX3_56 5'aattccctggag(n_{13})cg

EX3_57 5'aattccctggag(n_{13})ct

EX3_58 5'aattccctggag(n_{13})ga

EX3_59 5'aattccctggag(n_{13})gc

EX3_60 5'aattccctggag(n_{13})gg

EX3_61 5'aattccctggag(n_{13})gt

EX3_62 5'aattccctggag(n_{13})ta

EX3_63 5'aattccctggag(n_{13})tc

EX3_64 5'aattccctggag(n_{13})tg

EX3_65 5'aattccctggag(n_{13})tt

and their complementary strand (SEQ ID NO: 145):

EX3_66 : 5' (n_{13})ctccaggg

The reading frame is maintained as above and the string of 13 n's is to ensure that the Bpm1 site cuts beyond the capture dinucleotides one base into any captured fragment. All 16 possible vectors were produced. The complementary strand produces a 5' aatt overhang and the appropriate 3' overhang.

The series of vectors pREPLACEXnn were produced by oligonucleotides EX3_67 to EX3_130 (SEQ ID NOs: 146-209) replacing the EcoR1 to BsrD1 region of pINDaa to pINDtt. The first four oligonucleotides were for pINDaa, the second four for pINDac and so on until the end of the series. Respectively:

EX3_67 5'aattctggag(n_{14})aaa

EX3_68 5'aattctggag(n_{14})aac

EX3_69 5'aattctggag(n_{14})aag

EX3_70 5'aattctggag(n_{14})aat

EX3_71	5'aattctggag(n_{14})aca
EX3_72	5'aattctggag(n_{14})acc
EX3_73	5'aattctggag(n_{14})acg
EX3_74	5'aattctggag(n_{14})act
EX3_75	5'aattctggag(n_{14})aga
EX3_76	5'aattctggag(n_{14})agc
EX3_77	5'aattctggag(n_{14})agg
EX3_78	5'aattctggag(n_{14})agt
EX3_79	5'aattctggag(n_{14})ata
EX3_80	5'aattctggag(n_{14})atc
EX3_81	5'aattctggag(n_{14})atg
EX3_82	5'aattctggag(n_{14})att
EX3_83	5'aattctggag(n_{14})caa
EX3_84	5'aattctggag(n_{14})cac
EX3_85	5'aattctggag(n_{14})cag
EX3_86	5'aattctggag(n_{14})cat
EX3_87	5'aattctggag(n_{14})cca
EX3_88	5'aattctggag(n_{14})ccc
EX3_89	5'aattctggag(n_{14})ccg
EX3_90	5'aattctggag(n_{14})cct
EX3_91	5'aattctggag(n_{14})cga
EX3_92	5'aattctggag(n_{14})cgc
EX3_93	5'aattctggag(n_{14})cgg
EX3_94	5'aattctggag(n_{14})cgt
EX3_95	5'aattctggag(n_{14})cta
EX3_96	5'aattctggag(n_{14})ctc
EX3_97	5'aattctggag(n_{14})ctg
EX3_98	5'aattctggag(n_{14})ctt

EX3_99 5'aattctggag(n_{14})gaa

EX3_100 5'aattctggag(n_{14})gac

EX3_101 5'aattctggag(n_{14})gag

EX3_102 5'aattctggag(n_{14})gat

EX3_103 5'aattctggag(n_{14})gca

EX3_104 5'aattctggag(n_{14})gcc

EX3_105 5'aattctggag(n_{14})gcg

EX3_106 5'aattctggag(n_{14})gct

EX3_107 5'aattctggag(n_{14})gga

EX3_108 5'aattctggag(n_{14})ggc

EX3_109 5'aattctggag(n_{14})ggg

EX3_110 5'aattctggag(n_{14})ggt

EX3_111 5'aattctggag(n_{14})gta

EX3_112 5'aattctggag(n_{14})gtc

EX3_113 5'aattctggag(n_{14})gtg

EX3_114 5'aattctggag(n_{14})gtt

EX3_115 5'aattctggag(n_{14})taa

EX3_116 5'aattctggag(n_{14})tac

EX3_117 5'aattctggag(n_{14})tag

EX3_118 5'aattctggag(n_{14})tat

EX3_119 5'aattctggag(n_{14})tca

EX3_120 5'aattctggag(n_{14})tcc

EX3_121 5'aattctggag(n_{14})tcg

EX3_122 5'aattctggag(n_{14})tct

EX3_123 5'aattctggag(n_{14})tga

EX3_124 5'aattctggag(n_{14})tgc

EX3_125 5'aattctggag(n_{14})tgg

EX3_126 5'aattctggag(n_{14})tgt

EX3_127 5'aattctggag(n_{14})tta

EX3_128 5'aattctggag(n_{14})ttc

EX3_129 5'aattctggag(n_{14})ttg

EX3_130 5'aattctggag(n_{14})ttt

The complementary strands were EX3_131 to EX3_148 (SEQ ID NOs: 210-227):

EX3_131 5' t(n_{14})ctccag

EX3_132 5' g(n_{14})ctccag

EX3_133 5' c(n_{14})ctccag

EX3_134 5' a(n_{14})ctccag

EX_131 was used with EX3_67 to EX3_82, EX_132 with EX3_83 to EX3_98 and so on to the end of the series. The string of 14 n's is to ensure that the Bpm1 site cuts one base before the capture dinucleotides leaving an extra base on any captured fragment. All 64 possible vectors were produced so that any base could be added to any possible capture sequence. The complementary strands produced a 5' aatt overhang and the appropriate 3' overhang. Blue white colour selection was not maintained.

Cutting the appropriate pINDnn vector with BamH1 and BsrD1 and replacing with oligonucleotides below :

EX3_135 pFRAG16 5' gaggctcagt gatacagtct tccacggccg ttgtaaattg tcgggaagac tgctcctcca gcag

EX3_136 pFRAG15 5' gaggctcagg atacagtctt ctcacggccg ttgtaaattg tcggaagact gctccctcca gcag

EX3_137 pFRAG14 5' gaggctcaga tacagtcttc gtcacggccg ttgtaaattg tcgaagactg ctccgctcca gcag

EX3_138 pFRAG13 5' gaggctcgat acagtcttca gtcacggccg ttgtaaattg tgaagactgc tcccgctcca gcag

EX3_139 pFRAG12 5' gaggctgata cagtcttcca gtcacggccg ttgtaaattg gaagactgct cctcgctcca gcag

EX3_140 pFRAG11 5' gaggcgatac agtcttctca gtcacggccg ttgtaaattg aagactgctc cgtcgctcca gcag

EX3_141 pFRAG10 5' gagggataca gtcttcctca gtcacggccg ttgtaaatga agactgctcc tgtcgctcca gcag

EX3_142 pFRAG09 5' gaggatacag tcttcgctca gtcacggccg ttgtaaagaa gactgctcct tgtcgctcca gcag

EX3_143 pFRAG08 5' gagatacagt cttcggctca gtcacggccg ttgtaagaag actgctccat tgtcgctcca gcag

EX3_144 pFRAG07 5' ggatacagtc ttcaggctca gtcacggccg ttgtagaaga ctgctccaat tgtcgctcca gcag

EX3_145 pFRAG06 5' gatacagtct tcgaggctca gtcacggccg ttgtgaagac tgctccaaat tgtcgctcca gcag

EX3_146 pFRAG05 5' atacagtctt ctgaggctca gtcacggccg ttggaagact gctcctaaat tgtcgctcca gcag

EX3_147 pFRAG04 5' tacagtcttc gtgaggctca gtcacggccg ttgaagactg ctccgtaaat tgtcgctcca gcag

EX3_148 pFRAG03 5' acagtcttcc gtgaggctca gtcacggccg tgaagactgc tcctgtaaat tgtcgctcca gcag

with their complementary strands EX3_149 to EX3_162 (SEQ ID NOs: 228-241) where their complementary strands additionally had the BamH1 complementary end 5' gatc respectively :

EX3_149 pFRAG16

5'gatcctgctggaggagcagtcttcccgacaatttacaacggccgtggaagactgtatcactgagcctcac

EX3_150 pFRAG15

5'gatcctgctggagggagcagtcttccgacaatttacaacggccgtgagaagactgtatcctgagcctcac

EX3_151 pFRAG14

5'gatcctgctggagcggagcagtcttcgacaatttacaacggccgtgacgaagactgtatctgagcctcac

EX3_152 pFRAG13

5'gatcctgctggagcgggagcagtcttcacaatttacaacggccgtgactgaagactgtatcgagcctcac

EX3_153 pFRAG12

5'gatcctgctggagcgaggagcagtcttccaatttacaacggccgtgactggaagactgtatcagcctcac

EX3_154 pFRAG11

5'gatcctgctggagcgacggagcagtcttcaatttacaacggccgtgactgagaagactgtatcgcctcac

EX3_155 pFRAG10

5'gatcctgctggagcgacaggagcagtcttcatttacaacggccgtgactgaggaagactgtatccctcac

EX3_156 pFRAG09

5'gatcctgctggagcgacaaggagcagtcttctttacaacggccgtgactgagcgaagactgtatcctcac

EX3_157 pFRAG08

5'catcctgctggagcgacaatggagcagtcttcttacaacggccgtgactgagccgaagactgtatctcac

EX3_158 pFRAG07

5'gatcctgctggagcgacaattggagcagtcttctacaacggccgtgactgagcctgaagactgtatccac

EX3_159 pFRAG06

5'gatcctgctggagcgacaatttggagcagtcttcacaacggccgtgactgagcctcgaagactgtatcac

EX3_160 pFRAG05

5'gatcctgctggagcgacaatttaggagcagtcttccaacggccgtgactgagcctcagaagactgtatcc

EX3_161 pFRAG04

5'gatcctgctggagcgacaatttacggagcagtcttcaacggccgtgactgagcctcacgaagactgtatc

EX3_162 pFRAG03

5'gatcctgctggagcgacaatttacaggagcagtcttcacggccgtgactgagcctcacggaagactgtat

produced the series of vectors pFRAG16 to pFRAG03. Fourteen vectors were produced to clone PCR products to be fragmented from any possible starting point up to 14 bases in frame from each side. Note that pFRAG03 used pINDat, pFRAG04 used pINDga, pFRAG05 used pINDgg and the remainder used pINDgt.

p53 exon 5 was PCRd in a Mastercycler (Eppendorf) with primer pairs from the following sets.

Forward set 1 :

EX_163 5' gatactcaactctgtctccttcc, (SEQ ID NO: 242)

EX_164 5'gatactcaactctgtctccttcctcttc, (SEQ ID NO: 243)

EX_165 5' gatactcaactctgtctccttcctcttcctac, (SEQ ID NO: 244)

reverse set 1 :

EX_166 5' ggagccccacagctgcacagggca, (SEQ ID NO: 245)

EX_167 5' ggagccccacagctgcacagggcaggtct, (SEQ ID NO: 246)

EX_168 5' ggagccccacagctgcacagggcaggtcttg, (SEQ ID NO: 247)

forward set 2 :

EX_169 5' gatacgtgcagctgtgggttgattc, (SEQ ID NO: 248)

EX_170 5' gatacgtgcagctgtgggttgattccac, (SEQ ID NO: 249)

EX_171 5' gatacgtgcagctgtgggttgattccacac, (SEQ ID NO: 250)

reverse set 2 :

EX_172 5' ggagccacaacctccgtcatgtg, (SEQ ID NO: 251)

EX_173 5' ggagccacaacctccgtcatgtgctgtgac (SEQ ID NO: 252)

EX_174 5' ggagccacaacctccgtcatgtgctg (SEQ ID NO: 253)

forward set 3 :

EX_175 5'gatacgacggaggttgtgaggcg, (SEQ ID NO: 254)

EX_176 5' gatacgacggaggttgtgaggcgct, (SEQ ID NO: 255)

EX_177 5' gatacgacggaggttgtgaggcgctgcccccac, (SEQ ID NO: 256)
reverse set 3 :
EX_178 5' ggagcggcaaccagccctgtcgt, (SEQ ID NO: 257)
EX_179 5' ggagcggcaaccagccctgtcgtctct, (SEQ ID NO: 258)
EX_180 5' ggagcggcaaccagccctgtcgtctctcca. (SEQ ID NO: 259)

Each possible forward primer from a set was used with each possible reverse primer from the same set. PCR was performed at 94.5°C for 5 minutes, then 36 cycles of 94.5°C, 67°C and 72°C for 30 seconds each temperature and with a 5°C gradient at the 67° C step. A final incubation of 72 C for 10 minutes was performed. 50ul reactions containing 0.2mM dNTPs, 25 pmoles of each primer, 2.5 units of AmpliTaq Gold (Perkin Elmer) 2.5mM $MgCl_2$ and 10 to 100ng of human genomic DNA were used. All but 3 combinations gave the expected products demonstrating the utility of the approach. Note that the forward and reverse primers produce the ends required for directional cloning as described. The length of the PCR products was 83 bases, exactly as required for cutting the fragments at the same point from either end of the inserts excised from a given pFRAGnn vector. Together the three sets of PCR products were designed to cover the entire p53 exon 5 sequence. PCRs were purified in a QUIAquick 96 (Qiagen) and incubated for 30 minutes with 1 unit of T4 DNA polymerase (New England Biolabs) in the presence of 0.2mM dCTP for 37°C in a 50 ul reaction volume to expose the four 5' bases at each end for ligation to the pFRAGnn vectors. Reactions were purified in a QUIAspin column (Qiagen) and the resultant fragments cloned as described above into the ends produced by cutting the pFRAG03-16 vectors to completion with Bbs1 (New England Biolabs). The vectors were counter selected where possible by cutting ligations with 20 units of Eag1. White colonies were picked and the presence of the exact p53

sequence flanked by the primers was confirmed by sequencing as above. Correct plasmids were purified using QIA filter Plasmid maxipreps (Qiagen).

Cyclical cutting and ligation was performed as described in example 2 except that the clones were first cut to completion with Bpm1, EcoR1 and BamH1, and purified by QUIAquick columns(Qiagen). The adapter prevents cyclical cutting and ligation of the vector.

Fragments produced (cut insert plus ligated adapter) were separated through 15 % polyacrylamide electrophoresis gels, visualised using a fluorimager (Molecular Dynamics) excised and purified using QUIAEX II (Qiagen).

Each of the pSELECTnn vectors was ligated separately to purified fragments so that the dinucleotides at the ends of the captured fragments could be determined through knowing the capture dinucleotides. Ligation was performed in a 30 molar excess of blocking adapters and vector over fragments to prevent fragments from autoligating. Blocking adaptors had all possible 2 base 3' overhangs except the one that corresponded to the vector and its complement. Vector and vector plus captured fragments were purified from the remaining reaction components either by phenol/chloroform extraction followed by Sepharose size exclusion chromatograhy or by agarose gel electrophoresis followed by QIAEX II (Qiagen). Fragments were removed from the vector by cutting to completion with Bpm1 and purified as before. They were now ready for use for use as unlabeled fragments of known end sequence, known registration and corresponding exactly to the original sequence. Note that cutting occurs from the vector. We had found that following their purification if adapters were originally unphosphorylated at their 5' they were difficult to cut using Bpm1 sites in the adapters. The adapters lacked a 5' phosphate to

minimise their effectiveness on eachother during cyclical cutting and ligation. The unligated strand of the adapter can therefore be lost on purification thus inactivating any restriction sites that it contains.

The fragments were manipulated similarly with the pRESECTnn vectors with the aim of removing their exposed end base for later replacement. In this case, the vectors could optionally be pooled to select all fragments which ended with a particular base at the final 3' single stranded nucleotide. Ends not having the required sequence were blocked by appropriate pools of the blocking adaptors. Pooling pRESECTaa, pRESECTac, pRESECTag and pRESECTat for example captures all fragments which end with a t. Bpm1 releases the captured fragments and removes the t at their ends.

The pREPLACEXnn vectors were used as described above except that they captured fragments produced by the action of the pRESECTnn vectors. pREPLACEXnn vectors were used in pools. This time the pools were designed to add back a particular base or set of bases. If t had been removed as described above, pREPLACEtaa, tac, tag etc. to pREPLACEttt (16 in all) could be pooled to capture all possible resultant fragments. Cutting with Bpm1 in this case adds an a back to the fragments. Substituting the series: pREPLACEgaa, gac, gag etc. to pREPLACEgtt (16 in all) replaces the t with a g. The series of all plasmids except pREPLACEaaa, aac, aag etc. to pREPLACEatt (48 in all) produces populations of fragments in which every possible replacement of the t occurs. Blocking adapters were again used against unwanted ends. A fluorescent dUTP (R110 – Perkin Elmer) at 10 um was used to label the fragments attached to the. 10ug of vector plus fragments were used in a 20ul containing 1 unit of AmpliTaq gold (Perkin Elmer), 2.5 mM Mg^{2+}, 0.05mM of dCTP and ddGTP at 72°C for 30 minutes. This takes advantage of the adapter

lacking 5' phosphates so that one strand remains unligated to the fragment. The ligated strand acts as a template for adding label by polymerisation to the attached fragment. Labeling occurs opposite the a in the Bpm1 site and extension is limited to the next three bases as dATP and dGTP which would then be required are absent and ddGTP terminates the reaction at the first available c.

Detection of base differences was performed as described in the earlier examples. The quantities corresponded to those of example 1. p53 exon 5 was amplified separately from different samples of normal human DNA and from DNA isolated from human glioma or ovarian cancer cells. Amplification used the primers described in example 2. 50 ul reactions were performed containing 0.2mM dNTPs, 25 pmoles of each primer, 2.5 units of AmpliTaq Gold (Perkin Elmer) 2.5mM $MgCl_2$ and 1 to 10ng of human genomic DNA. 10% of the dTTP was replaced by biotinylated dUTP. PCR was performed at 94.5°C for 5 minutes, then 40 cycles of 94.5°C, 63°C and 72°C for 30 seconds each. A final incubation of 72°C for 10 minutes was performed. Fragments of an insert were produced from a given pFRAGnn vector and then further processed either by a pSELECTnn vector or the corresponding pRESECTnn and pREPLACEnn vectors. Resultant fragments from either process were then pooled according to the intended screen, heated to 95°C for 5 minutes and added to each separate sample of p53 exon 5 in 20 ul ligation reactions (New England Biolabs). Annealing was allowed at 37°C for 30 minutes and then 0.2 units of T4 DNA ligase were added. Ligation was allowed to proceed for 16 hours. Reactions were bound to streptavidin coated beads (Dynal). Bound material was washed as recommended and analysed by denaturing polyacrylamide gel electrophoresis using the ABI377 (Perkin Elmer) as recommended except that on occasion gels of 10 or 15 % polyacrylamide were used to increase separation in the size range

of interest. Ligation is scored if fragments corresponding to a labeled fragment plus the juxtaposed fragment attached to the adaptor are observed. Base substitutions are detected if fragments from a given pSELECTnn vector are able to ligate in a template dependent fashion to fragments from the combined action of pRESECTnn and pREPLACEnn vectors both sets of fragments having originated from a given pFRAGnn vector. The pSELECT vector indicates the dinucleotides next to the point of ligation. The pRESECT and pREPLACEXnn vectors indicate the nature of the substitution. Note that fragments that have been labeled and substituted can only ligate in a template dependent fashion if there is a corresponding substitution in the target. Such fragments are unable to ligate to eachother because they have a terminal dideoxynucleotide at their 3' end. They are therefore dependent for ligation on the fragments from the pSELECTnn vectors. The latter are dependent on the former for becoming labeled on ligation.

Similarly, use of fragments originally from 2 different pFRAGnn vectors that have been labeled as described for the pREPLACEXnn vectors but originate entirely from a pSELECTnn vector can indicate a deletion or a substitution. Successful ligation between fragments of a pFRAG03 with fragments of a pFRAG05 corresponds to a 2 base deletion. The dinuceotides adjacent to the deletion are indicated by the particular pSELECTnn vector which gave rise to the successful ligations. Producing fragments from a pREPLACEXnn vector without first using the pRESECTnn vectors to remove a base results in a base insertion. Such fragments used with pSELECTnn fragments produced from the same pFRAGnn vector can detect the corresponding insertion.

The target is used to capture the ligation products because 3' single stranded dinucleotides can form if fragments reanneal thus allowing fragments with

complementary ends to join. This is target independent ligation. Capturing the target together with fragments that are annealed to it as a result of template dependent ligation assures that the only ligations that are detected are the latter.

CLAIMS

1. A method of comparing first and second sample polynucleotides, comprising the steps of:

i) providing at least two different sub-populations of the first sample, each sub-population comprising a series of fragments of the first sample polynucleotide of known length, the 5' terminus of each fragment being located at a known position in the first sample polynucleotide;

ii) with each of the first sample sub-populations, providing a plurality of modification libraries by dividing the sub-population into a set of modification libraries and modifying the nucleic acid at a fixed position or plurality of fixed positions in each fragment, the modification libraries of the sub-populations between them providing for modification at each position in the first sample polynucleotide.

iii) contacting each modification library of each sub-population with the second sample polynucleotide under stringent hybridisation conditions and detecting the hybridisation of the second sample polynucleotide to the fragments of each modification library; and

iv) correlating the results of detection step (iii) to determine any differences between the first and second polynucleotides.

2. A method according to claim 1, the fragments of each sub-population being of different lengths.

3. A method according to claim 1, the fragments of each sub-population being of the same length.

4. A method of comparing first and second sample polynucleotides according to any one of the preceding claims, the first and second sample polynucleotides being of unknown sequence.

5. A method according to any one of the preceding claims, each modification library having a different modification at said fixed position.

6. A method according to any one of the preceding claims, a series of fragments of known length being obtainable from a sub-population or combination of sub-populations such that the sub-population or sub-populations form a contiguous series of fragments of the first sample polynucleotide.

7. A method according to any one of the preceding claims, the modification to each sub-population at the fixed position or fixed positions being selected from the group of substitution, deletion and addition of a nucleotide, and inversion of a pair of nucleotides..

8. A method according to claim 7, the modification being substitution and each sub-population being divided into twelve modification libraries, between them providing for each possible substitution of each nucleic acid, each modification library providing for one substitution of one nucleic acid.

9. A method according to claim 7, the modification being substitution and each sub-population being divided into four modification libraries, between them

providing for substitution by the same nucleic acid of each nucleic acid of the first sample polynucleotide.

10. A method according to any one of the preceding claims, the modification of the nucleic acids of a modification library occurring at the 3' or 5' terminus of the fragments.

11. A method according to any one of the preceding claims, comprising prior to the step of contacting each modification library of each sub-population with the second sample polynucleotide, the additional step of labelling the fragments of each sub-population.

12. A method according to claim 11, the label being selected from the group of a mass label, a chemical label, a ligand, an enzyme and a radiolabel.

13. A method according to claim 12, the label being a chemical label comprising a coloured dye.

14. A method according to any one of claims 11-13, comprising labelling the 5' or 3' terminus of the fragments of each sub-population.

15. A method according to any one of claims 11-14, labelling being performed when nucleic acids are modified to form the modification libraries.

16. A method according to any one of the preceding claims, the step of contacting each modification library of each sub-population with the second sample polynucleotide being carried out in the presence of the sub-population.

17. A method according to claim 16, the fragments of each modification library being modified such that they have a 3' dideoxynucleotide, each modification library sub-population being labelled at the 5' terminus, the step of contacting each modification library of each sub-population with the second sample polynucleotide being performed in the presence of a ligase enzyme, and the detection of hybridisation comprising detecting ligation products.

18. A set of modification libraries of at least two different sub-populations of a sample polynucleotide, each sub-population comprising a series of fragments of the sample polynucleotide of equal length, the 5' terminus of each fragment being located at a position in the sample polynucleotide n nucleotides from each adjacent fragment wherein n is at least 2, the modification libraries comprising sub-divisions of each sub-population having a modified nucleic acid at a fixed position or plurality of fixed positions in each fragment, each modification library having a different modification at said fixed position, and the modification libraries of the sub-populations between them providing for modification at each position in the first sample polynucleotide.

19. A set of modification libraries according to claim 18, the fragments of each sub-population being of the same length.

20. A set of modification libraries according to either one of claims 18 or 19, at least one sub-population comprising a series of fragments of length n such that the population forms a contiguous series of fragments of the first sample polynucleotide.

21. A set of modification libraries according to any one of claims 18-20, the modification to each sub-population at the fixed position or fixed positions being selected from the group of substitution, deletion and addition of a nucleotide.

22. A set of modification libraries according to claim 21, the modification being substitution and each sub-population being divided into twelve modification libraries, between them providing for each possible substitution of each nucleic acid, each modification library providing for one substitution of one nucleic acid.

23. A set of modification libraries according to claim 21, the modification being substitution and each sub-population being divided into four modification libraries, between them providing for substitution by the same nucleic acid of each nucleic acid of the first sample polynucleotide.

24. A set of modification libraries according to any one of claims 18-23, the modification of the nucleic acids of a modification library being at the 3' terminus of the fragments.

25. A set of modification libraries according to any one of claims 18-24, the fragments of each sub-population being labelled.

26. A set of modification libraries according to claim 25, the label being selected from the group of a mass label, a chemical label, a ligand, an enzyme and a radiolabel.

27. A set of modification libraries according to claim 26, the label being a chemical label comprising a coloured dye.

28. A set of modification libraries according to any one of claims 25-27, the label being at the 5' terminus of the fragments of each sub-population.

29. The use of a set of modification libraries according to any one of claims 18-28 in a method according to any one of claims 1-17.

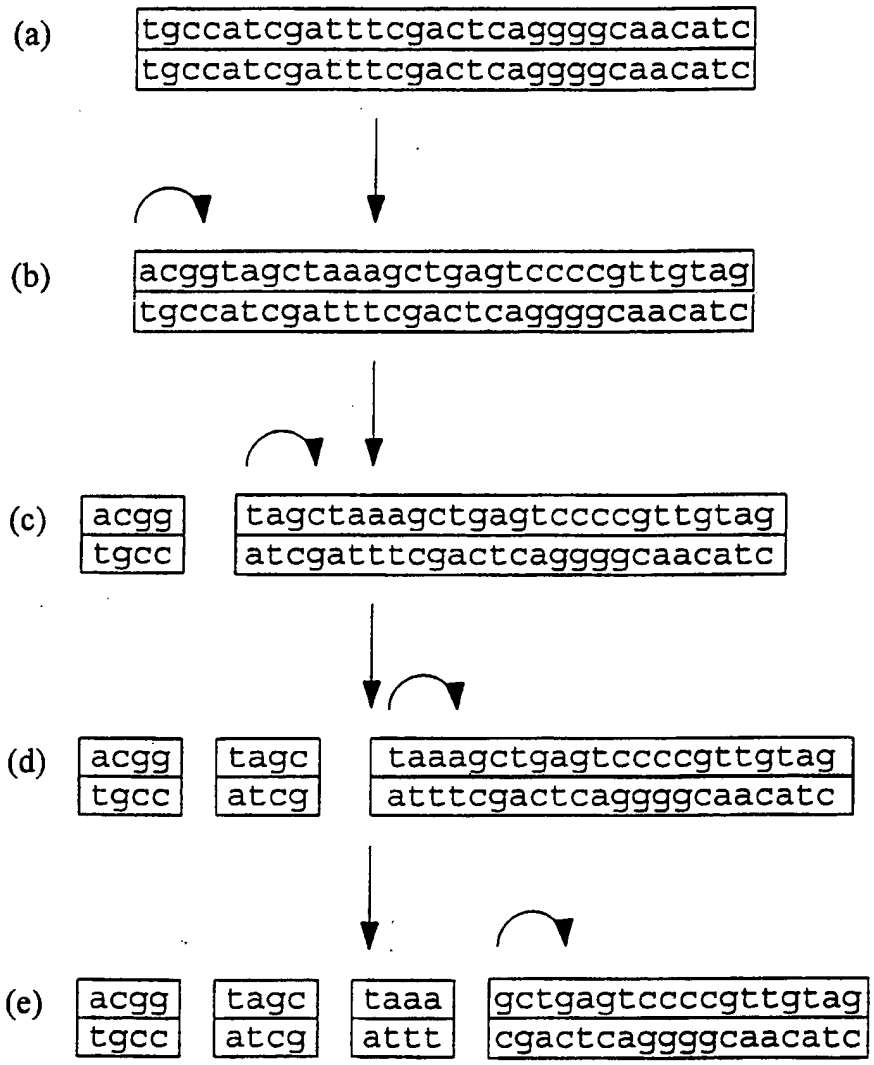

(a)
tgccatcgatttcgactcaggggcaacatc
tgccatcgatttcgactcaggggcaacatc
(b)
acggtagctaaagctgagtccccgttgtag
tgccatcgatttcgactcaggggcaacatc
(c)
acgg
tgcc
tagctaaagctgagtccccgttgtag
atcgatttcgactcaggggcaacatc
(d)
acgg
tgcc
tagc
atcg
taaagctgagtccccgttgtag
atttcgactcaggggcaacatc
(e)
acgg
tgcc
tagc
atcg
taaa
attt
gctgagtccccgttgtag
cgactcaggggcaacatc

Figure 1

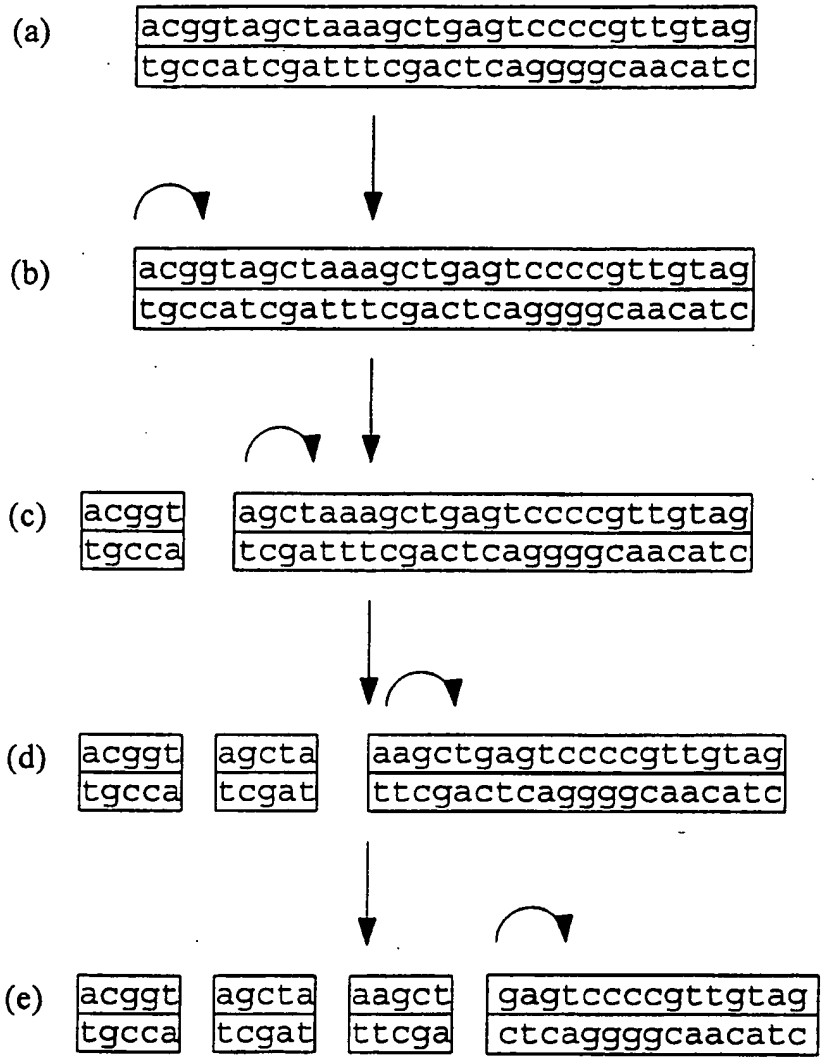

(a)
acggtagctaaagctgagtccccgttgtag
tgccatcgatttcgactcaggggcaacatc
(b)
acggtagctaaagctgagtccccgttgtag
tgccatcgatttcgactcaggggcaacatc
(c)
acggt
tgcca
agctaaagctgagtccccgttgtag
tcgatttcgactcaggggcaacatc
(d)
acggt
tgcca
agcta
tcgat
aagctgagtccccgttgtag
ttcgactcaggggcaacatc
(e)
acggt
tgcca
agcta
tcgat
aagct
ttcga
gagtccccgttgtag
ctcaggggcaacatc

Figure 2

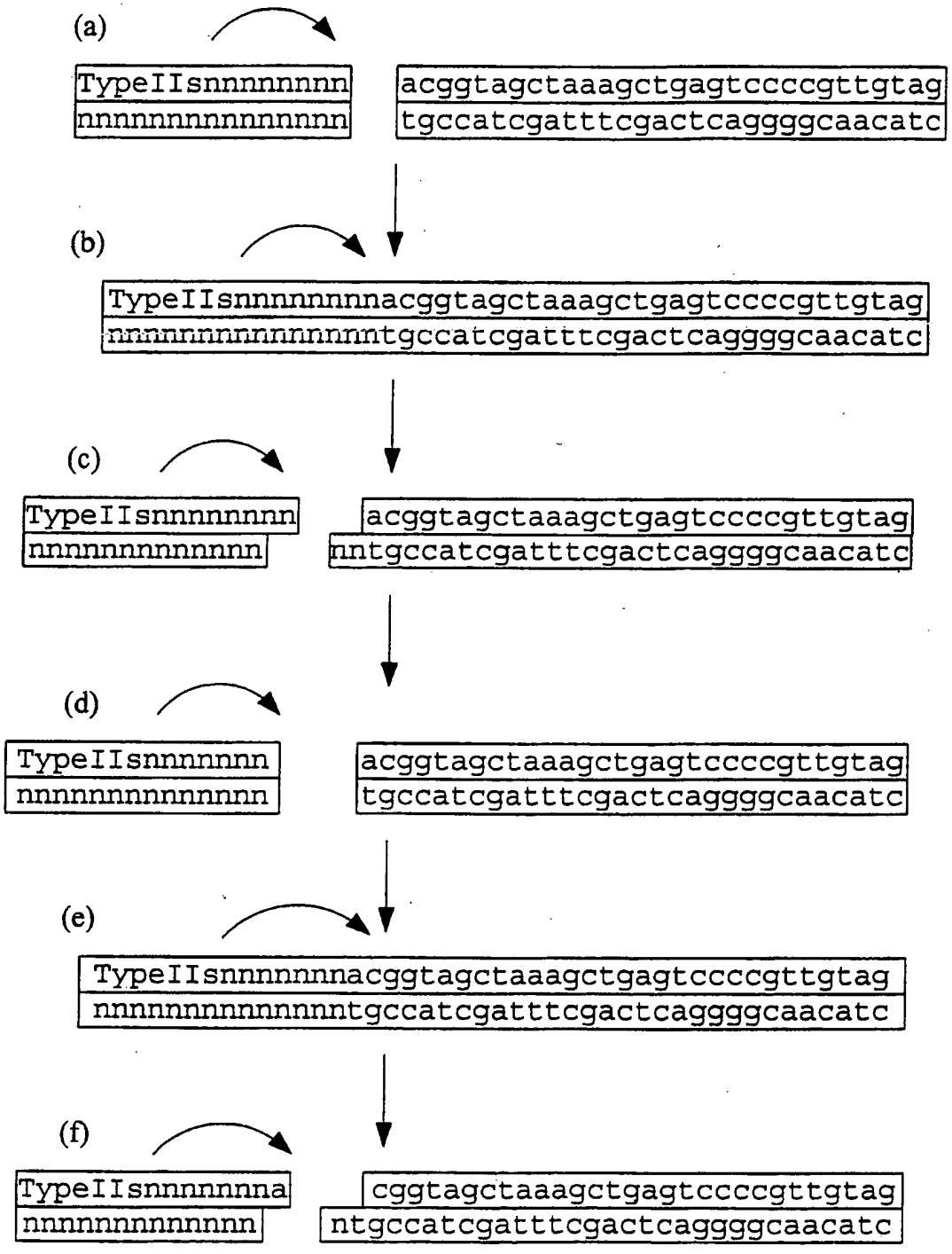

(a)
TypeIIsnnnnnnnn
nnnnnnnnnnnnnnn
acggtagctaaagctgagtccccgttgtag
tgccatcgatttcgactcaggggcaacatc
(b)
TypeIIsnnnnnnnnacggtagctaaagctgagtccccgttgtag
nnnnnnnnnnnnnnnntgccatcgatttcgactcaggggcaacatc
(c)
TypeIIsnnnnnnnn
nnnnnnnnnnnnn
acggtagctaaagctgagtccccgttgtag
nntgccatcgatttcgactcaggggcaacatc
(d)
TypeIIsnnnnnnn
nnnnnnnnnnnnnn
acggtagctaaagctgagtccccgttgtag
tgccatcgatttcgactcaggggcaacatc
(e)
TypeIIsnnnnnnnacggtagctaaagctgagtccccgttgtag
nnnnnnnnnnnnnnntgccatcgatttcgactcaggggcaacatc
(f)
TypeIIsnnnnnnna
nnnnnnnnnnnnn
cggtagctaaagctgagtccccgttgtag
ntgccatcgatttcgactcaggggcaacatc

Figure 3

Figure 3 (ctd.)

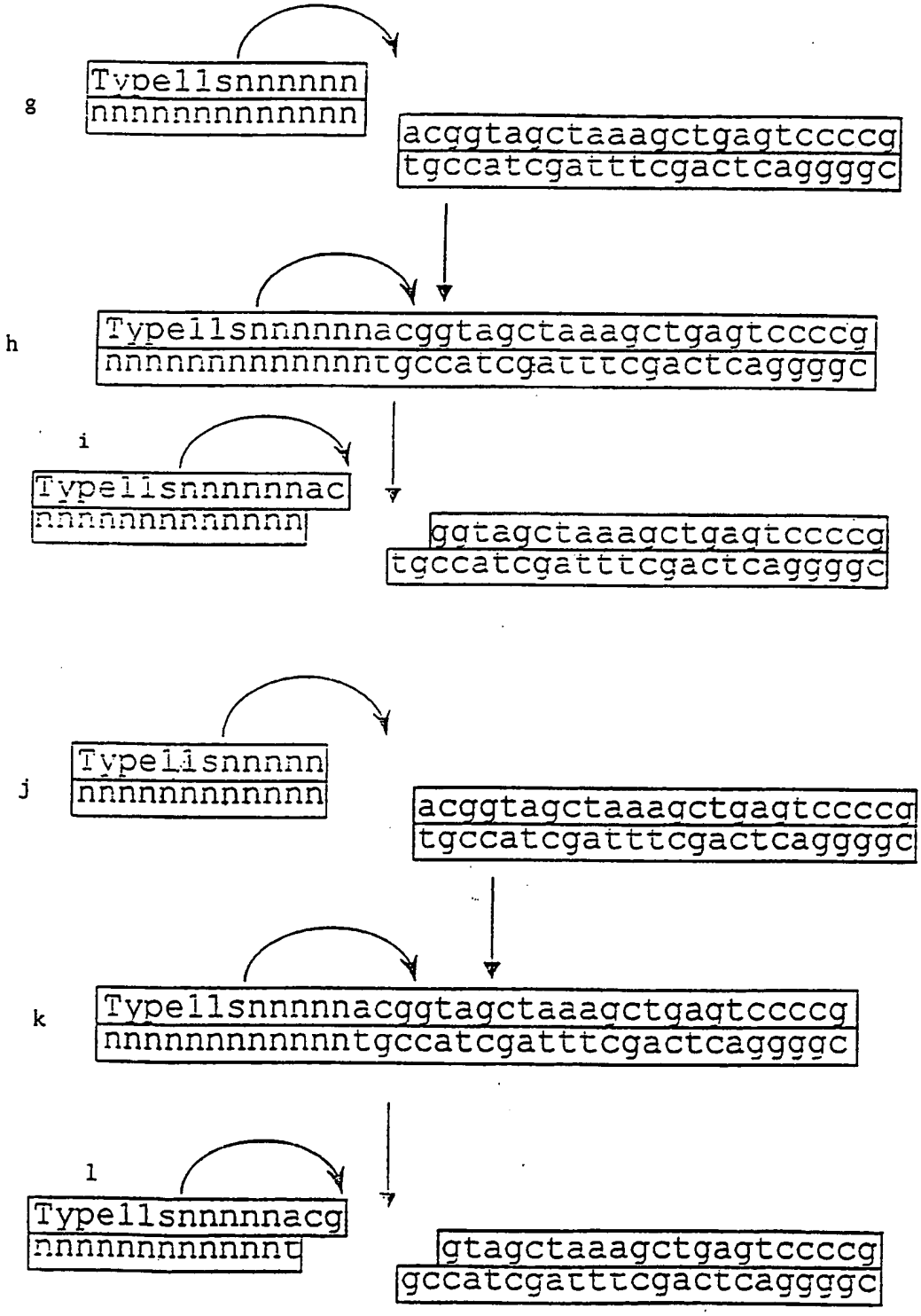

g
TypeIIsnnnnnn
nnnnnnnnnnnnn
acggtagctaaagctgagtccccg
tgccatcgatttcgactcaggggc
h
TypeIIsnnnnnnacggtagctaaagctgagtccccg
nnnnnnnnnnnnntgccatcgatttcgactcaggggc
i
TypeIIsnnnnnnac
nnnnnnnnnnnnn
ggtagctaaagctgagtccccg
tgccatcgatttcgactcaggggc
j
TypeIIsnnnnn
nnnnnnnnnnnn
acggtagctaaagctgagtccccg
tgccatcgatttcgactcaggggc
k
TypeIIsnnnnnacggtagctaaagctgagtccccg
nnnnnnnnnnnntgccatcgatttcgactcaggggc
l
TypeIIsnnnnnacg
nnnnnnnnnnnnt
gtagctaaagctgagtccccg
gccatcgatttcgactcaggggc

Figure 4

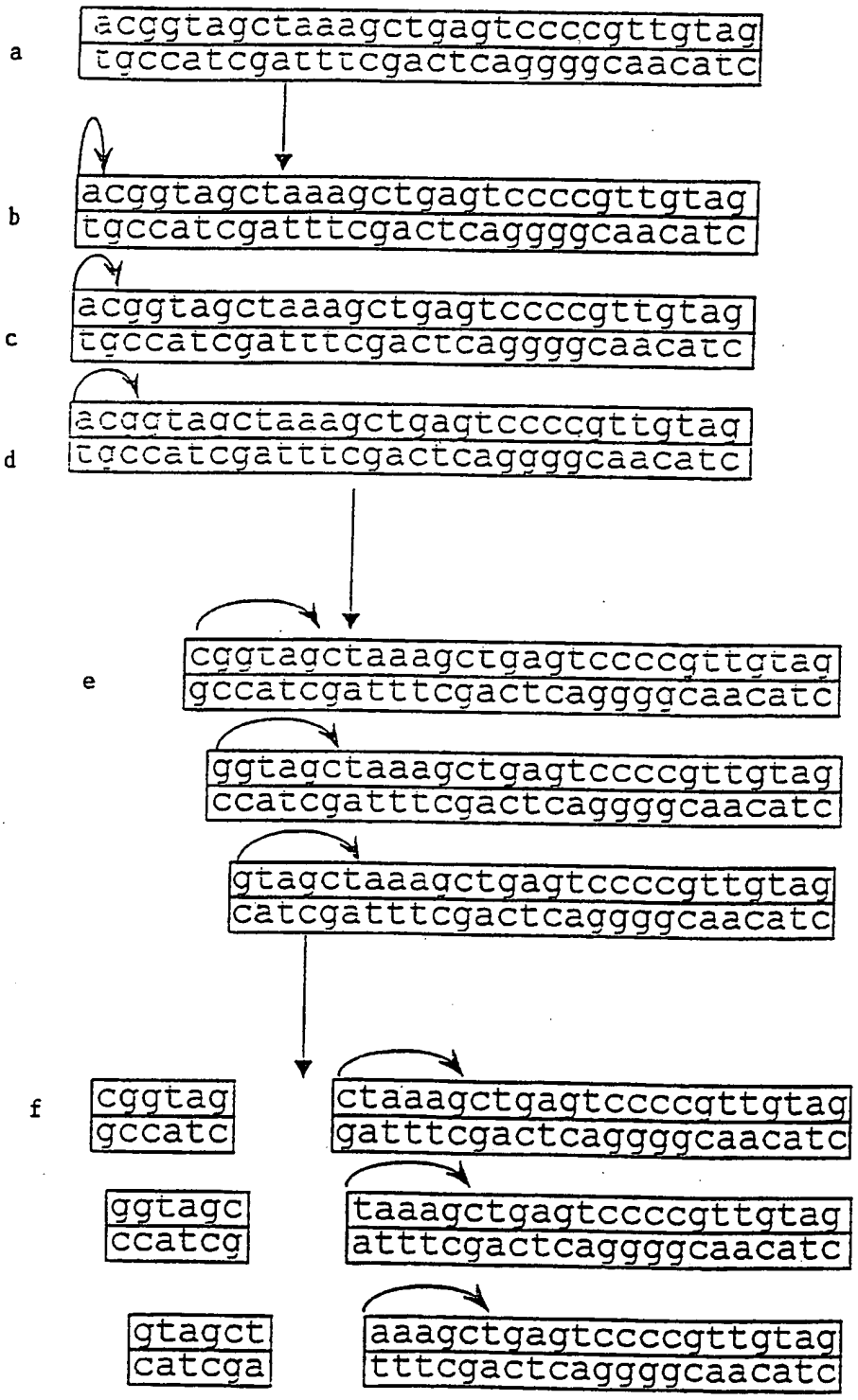

a
acggtagctaaagctgagtccccgttgtag
tgccatcgatttcgactcaggggcaacatc
b
acggtagctaaagctgagtccccgttgtag
tgccatcgatttcgactcaggggcaacatc
c
acggtagctaaagctgagtccccgttgtag
tgccatcgatttcgactcaggggcaacatc
d
acggtagctaaagctgagtccccgttgtag
tgccatcgatttcgactcaggggcaacatc
e
cggtagctaaagctgagtccccgttgtag
gccatcgatttcgactcaggggcaacatc
ggtagctaaagctgagtccccgttgtag
ccatcgatttcgactcaggggcaacatc
gtagctaaagctgagtccccgttgtag
catcgatttcgactcaggggcaacatc
f
cggtag
gccatc
ctaaagctgagtccccgttgtag
gatttcgactcaggggcaacatc
ggtagc
ccatcg
taaagctgagtccccgttgtag
atttcgactcaggggcaacatc
gtagct
catcga
aaagctgagtccccgttgtag
tttcgactcaggggcaacatc

Figure 5

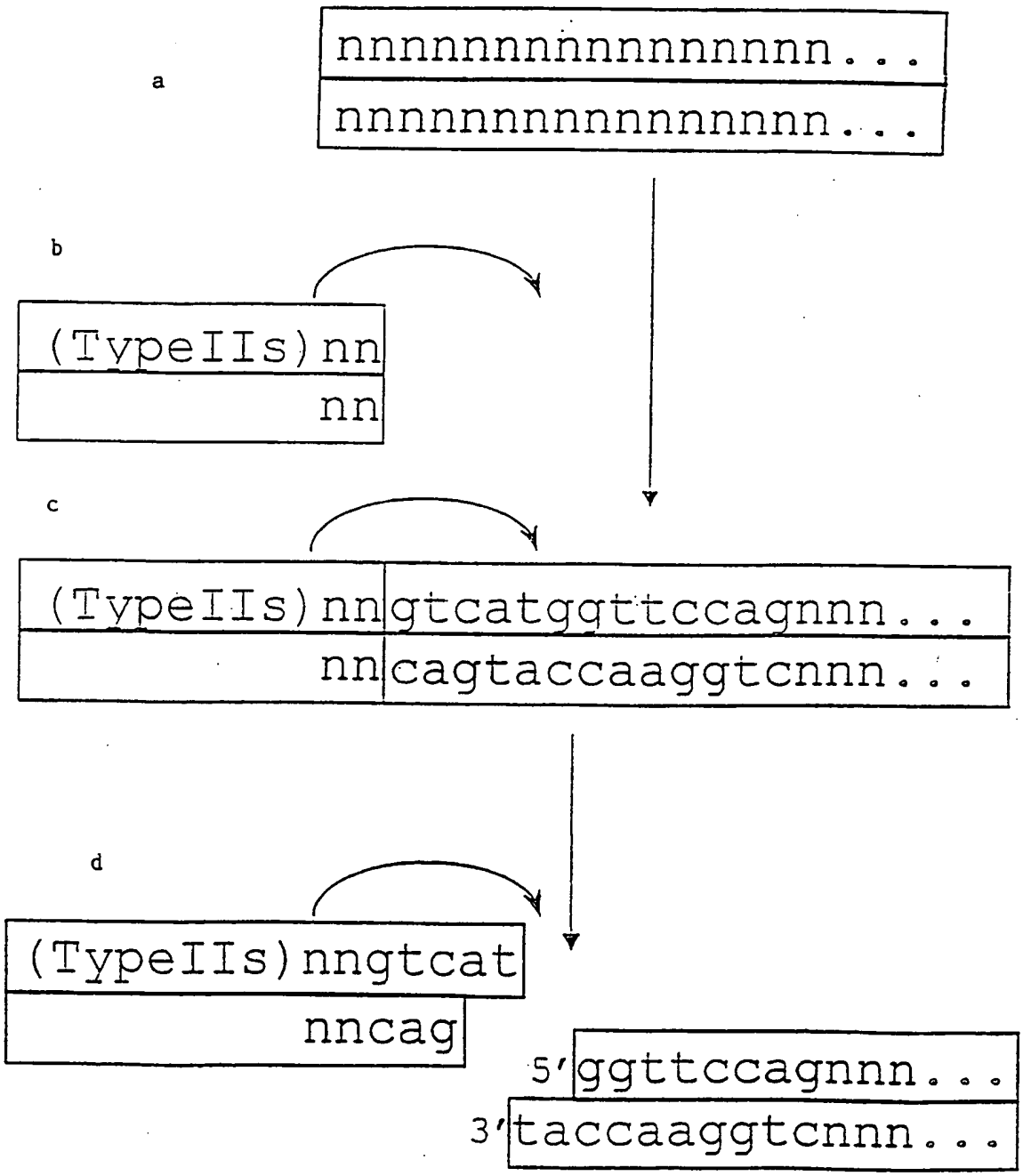

a
nnnnnnnnnnnnnnnn...
nnnnnnnnnnnnnnnn...
b
(TypeIIs)nn
nn
c
(TypeIIs)nngtcatggttccagnnn...
nncagtaccaaggtcnnn...
d
(TypeIIs)nngtcat
nncag
5'ggttccagnnn...
3'taccaaggtcnnn...

Figure 6

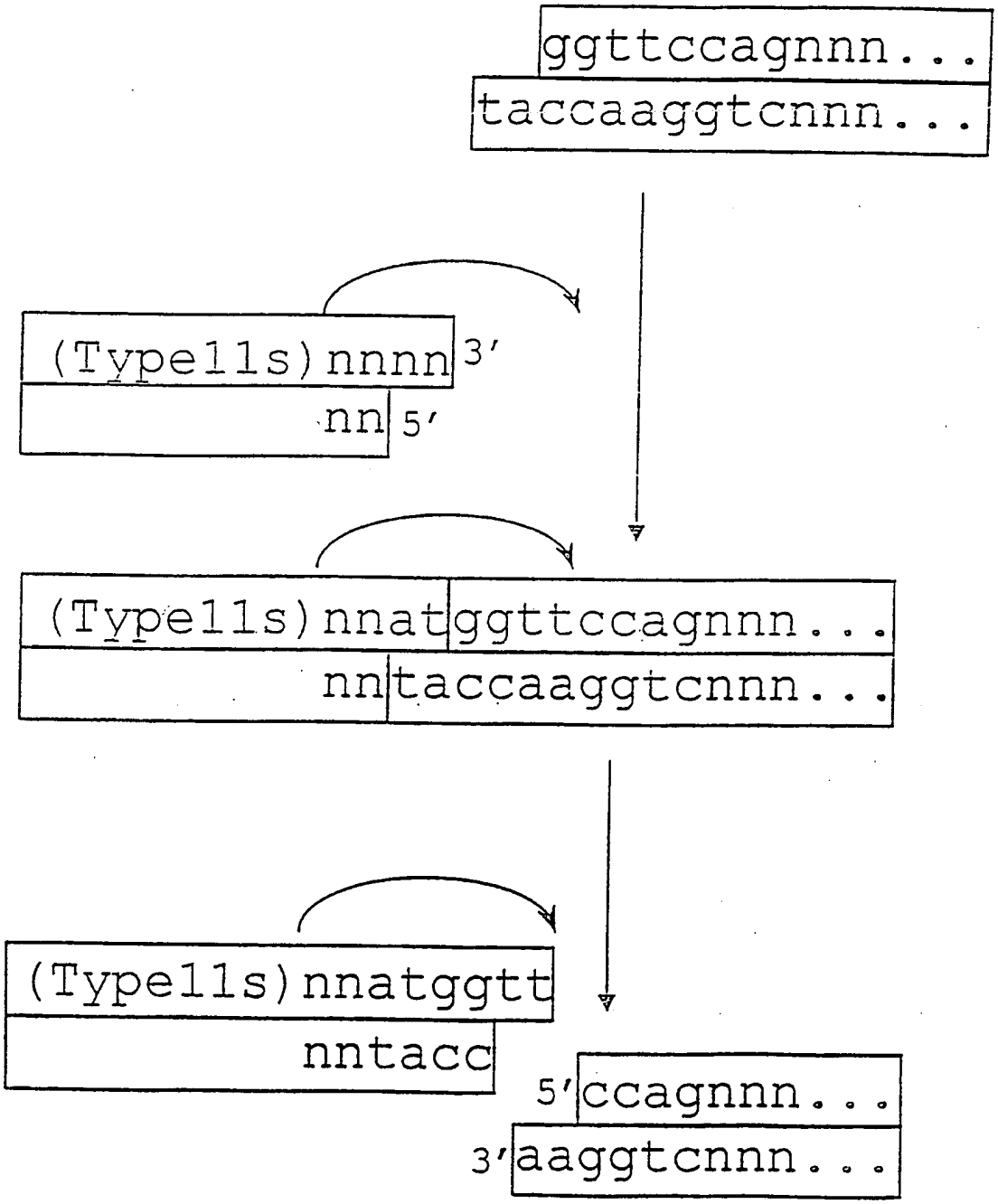

ggttccagnnn...
taccaaggtcnnn...
(TypeIIs)nnnn 3′
nn 5′
(TypeIIs)nnatggttccagnnn...
nntaccaaggtcnnn...
(TypeIIs)nnatggtt
nntacc
5′ccagnnn...
3′aaggtcnnn...

Figure 7

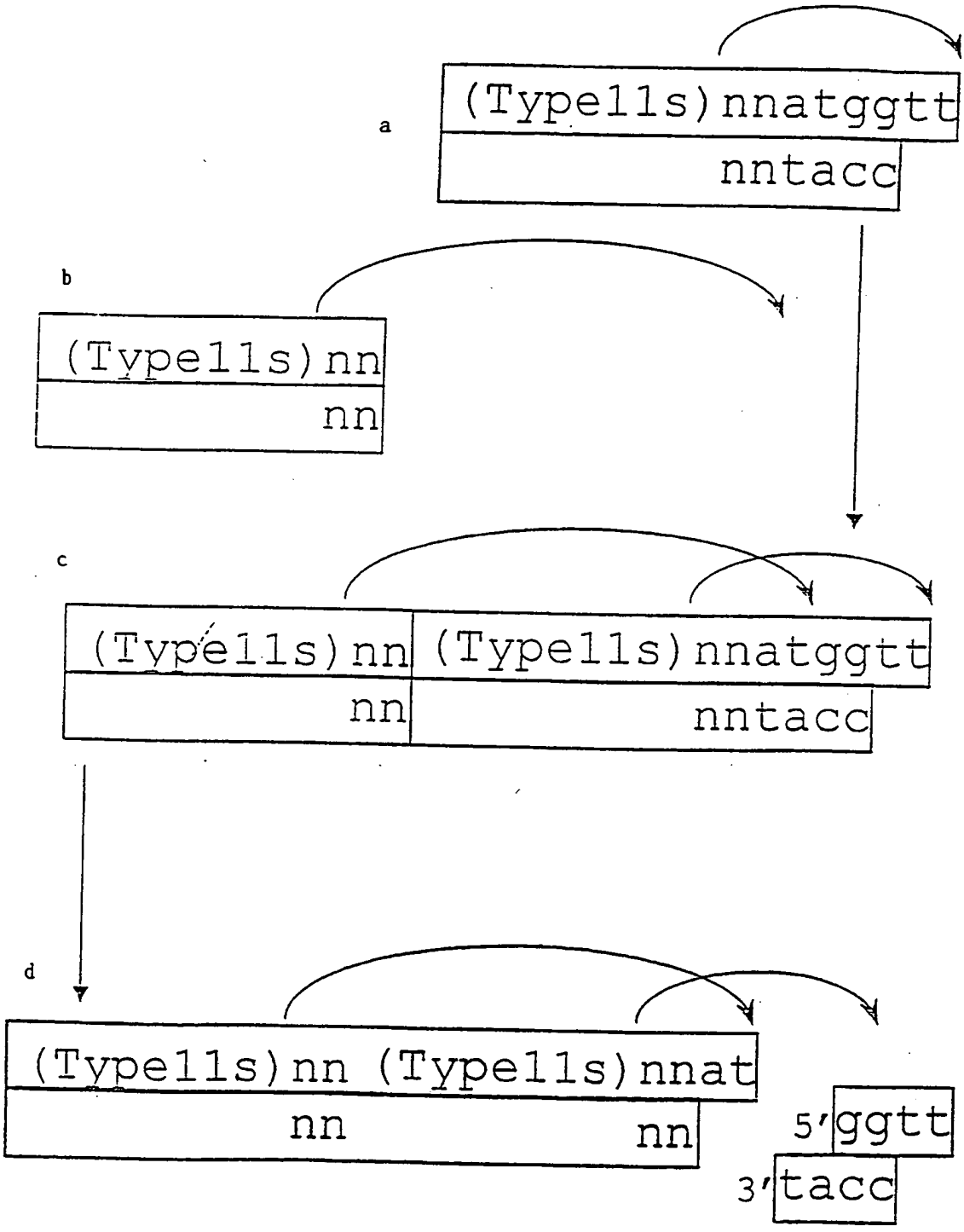

a
(Type11s)nnatggtt
nntacc
b
(Type11s)nn
nn
c
(Type11s)nn
nn
(Type11s)nnatggtt
nntacc
d
(Type11s)nn
nn
(Type11s)nnat
nn
5′ggtt
3′tacc

Figure 8

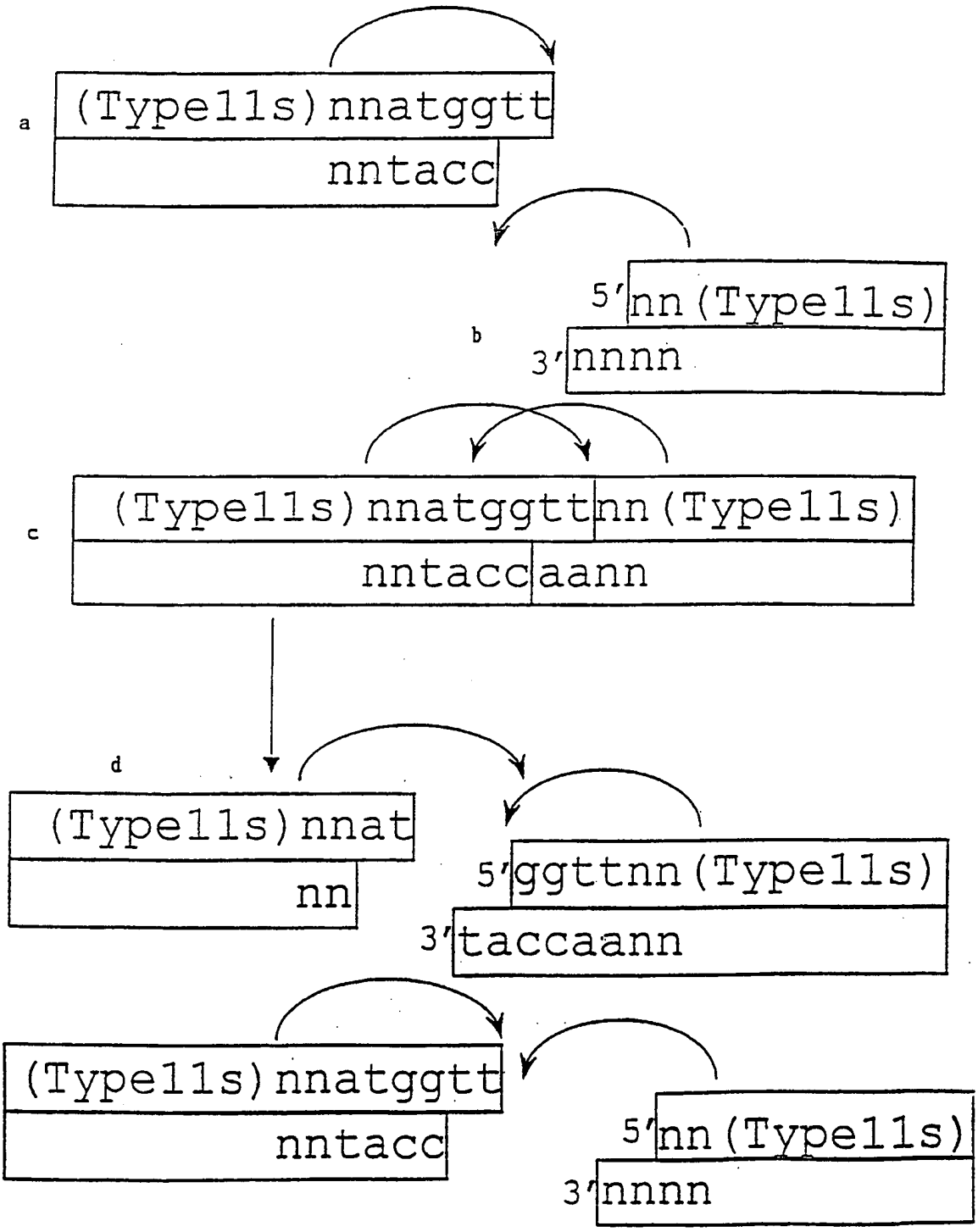

a
(Type11s)nnatggtt
nntacc
b
5′nn(Type11s)
3′nnnn
c
(Type11s)nnatggttnn(Type11s)
nntaccaann
d
(Type11s)nnat
nn
5′ggttnn(Type11s)
3′taccaann
(Type11s)nnatggtt
nntacc
5′nn(Type11s)
3′nnnn

Figure 9

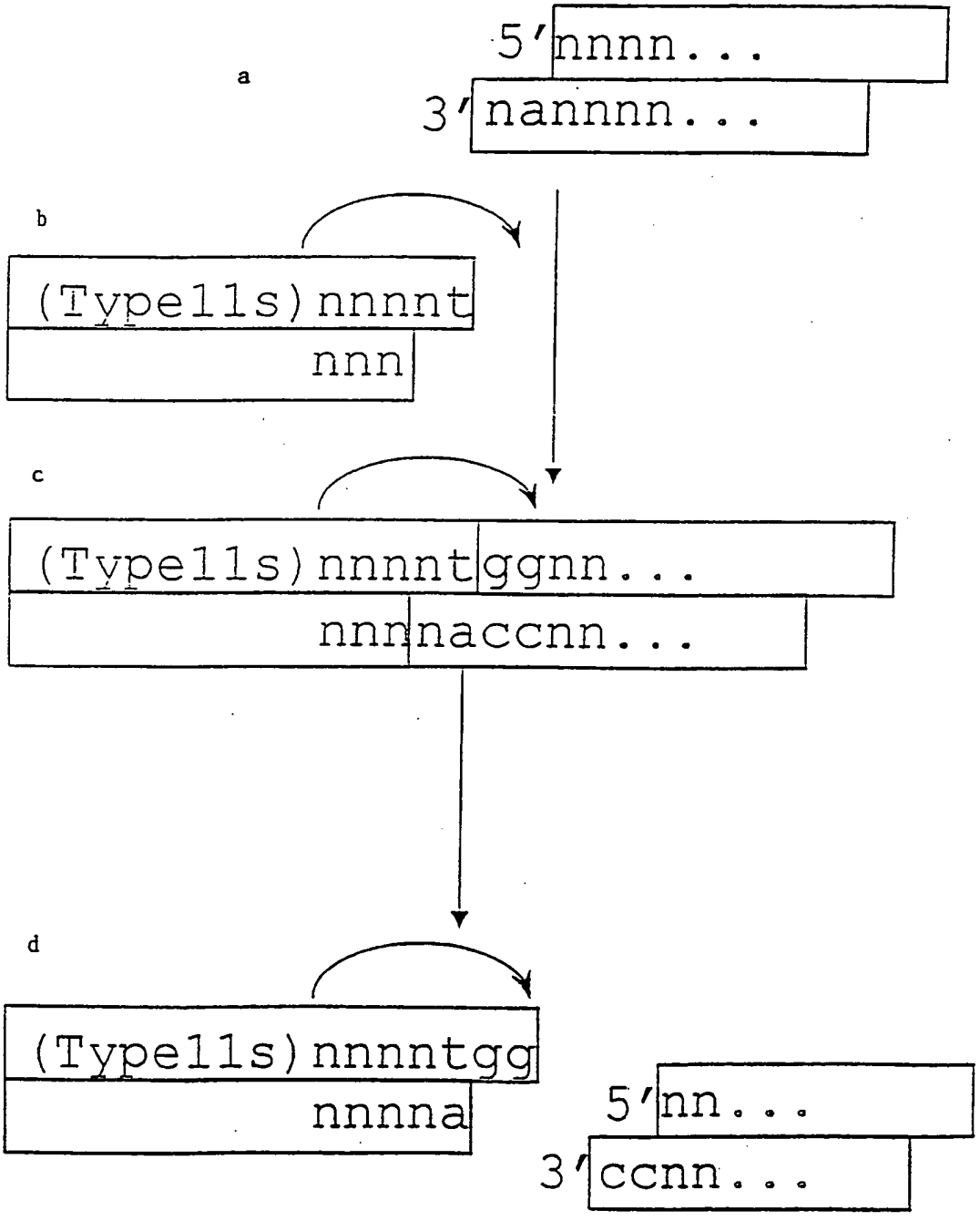

a
5′nnnn...
3′nannnn...
b
(Type11s)nnnnt
nnn
c
(Type11s)nnnntggnn...
nnnnaccnn...
d
(Type11s)nnnntgg
nnnna
5′nn...
3′ccnn...

Figure 10

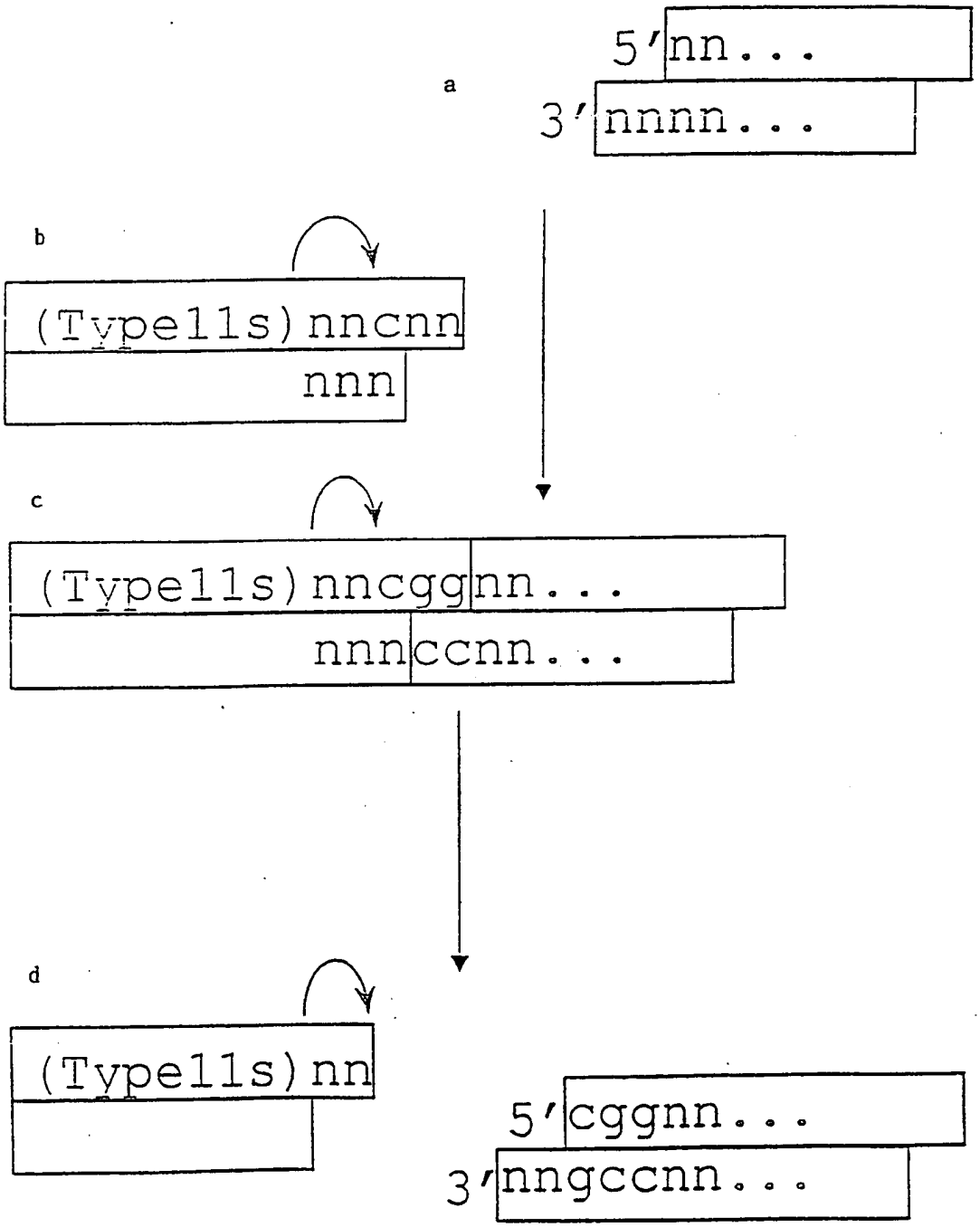

a
5′nn...
3′nnnn...
b
(Type11s)nncnn
nnn
c
(Type11s)nncggnn...
nnnccnn...
d
(Type11s)nn
5′cggnn...
3′nngccnn...

Figure 11

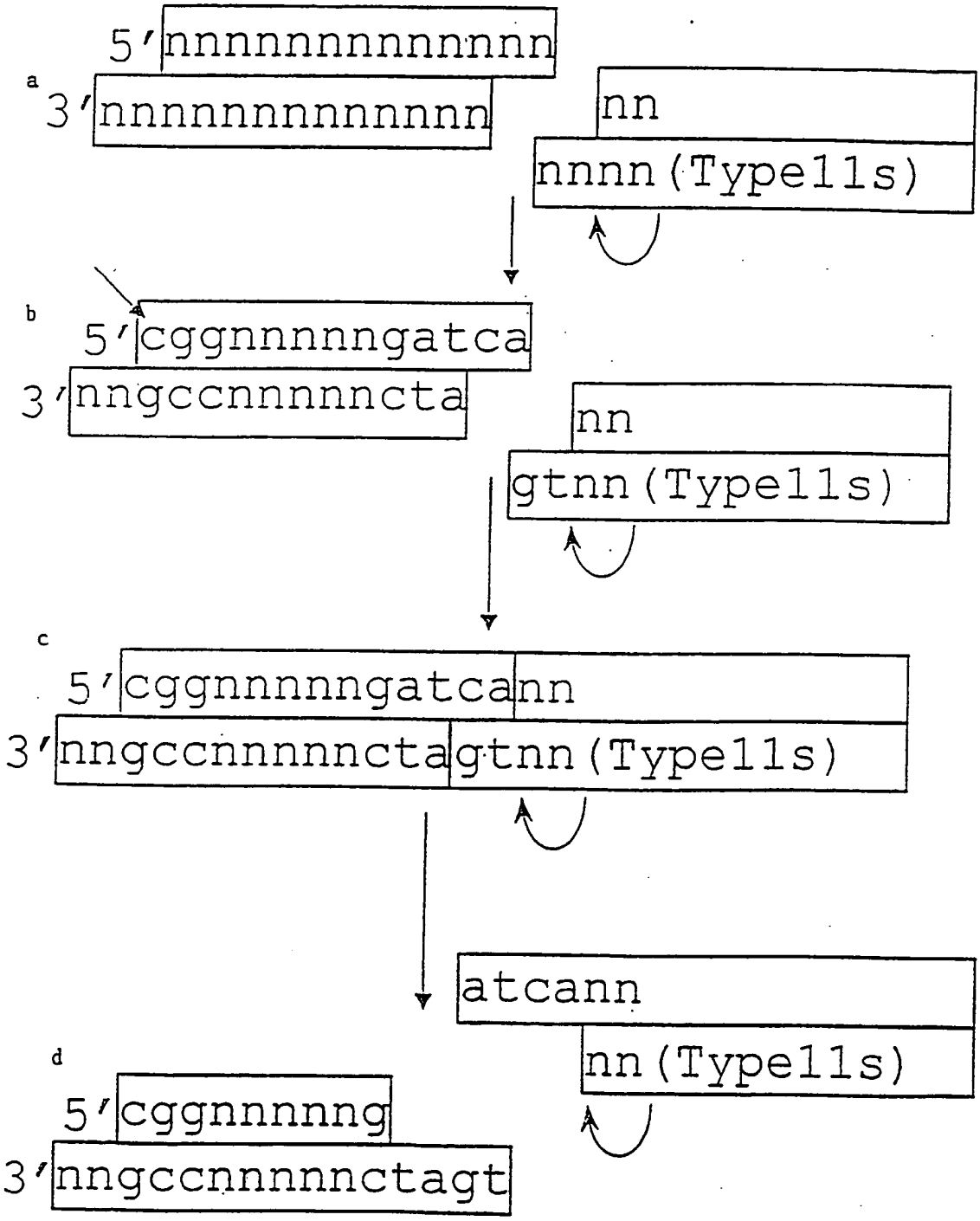

a
5′nnnnnnnnnnnnn
3′nnnnnnnnnnnnn
nn
nnnn(Type11s)
b
5′cggnnnnngatca
3′nngccnnnnnncta
nn
gtnn(Type11s)
c
5′cggnnnnngatcann
3′nngccnnnnnnctagtnn(Type11s)
atcann
nn(Type11s)
d
5′cggnnnnng
3′nngccnnnnnnctagt

Figure 12

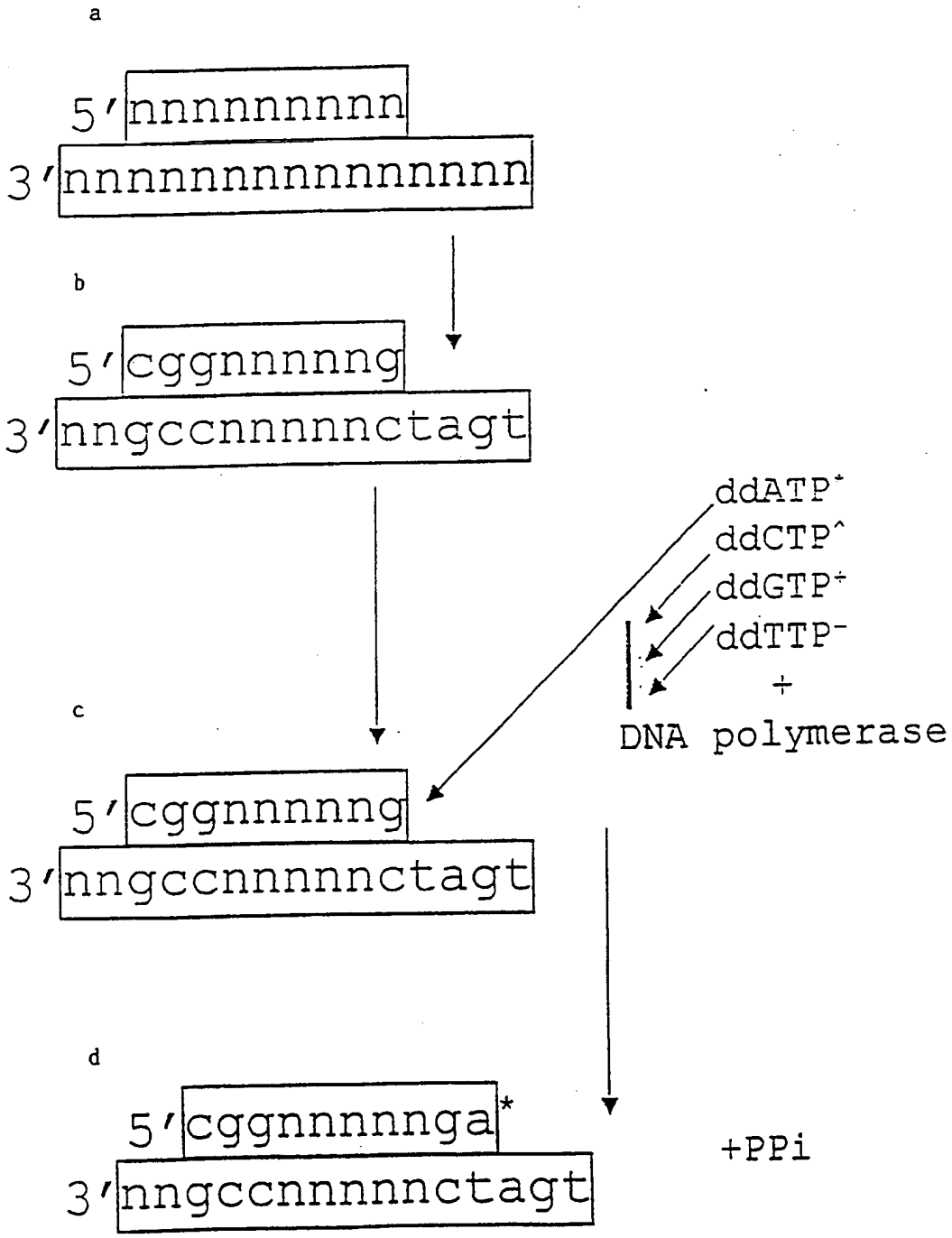

a
5′nnnnnnnnn
3′nnnnnnnnnnnnnnn
b
5′cggnnnnng
3′nngccnnnnnctagt
ddATP*
ddCTP^
ddGTP+
ddTTP-
+
DNA polymerase
c
5′cggnnnnng
3′nngccnnnnnctagt
d
5′cggnnnnnga*
3′nngccnnnnnctagt
+PPi

Figure 13

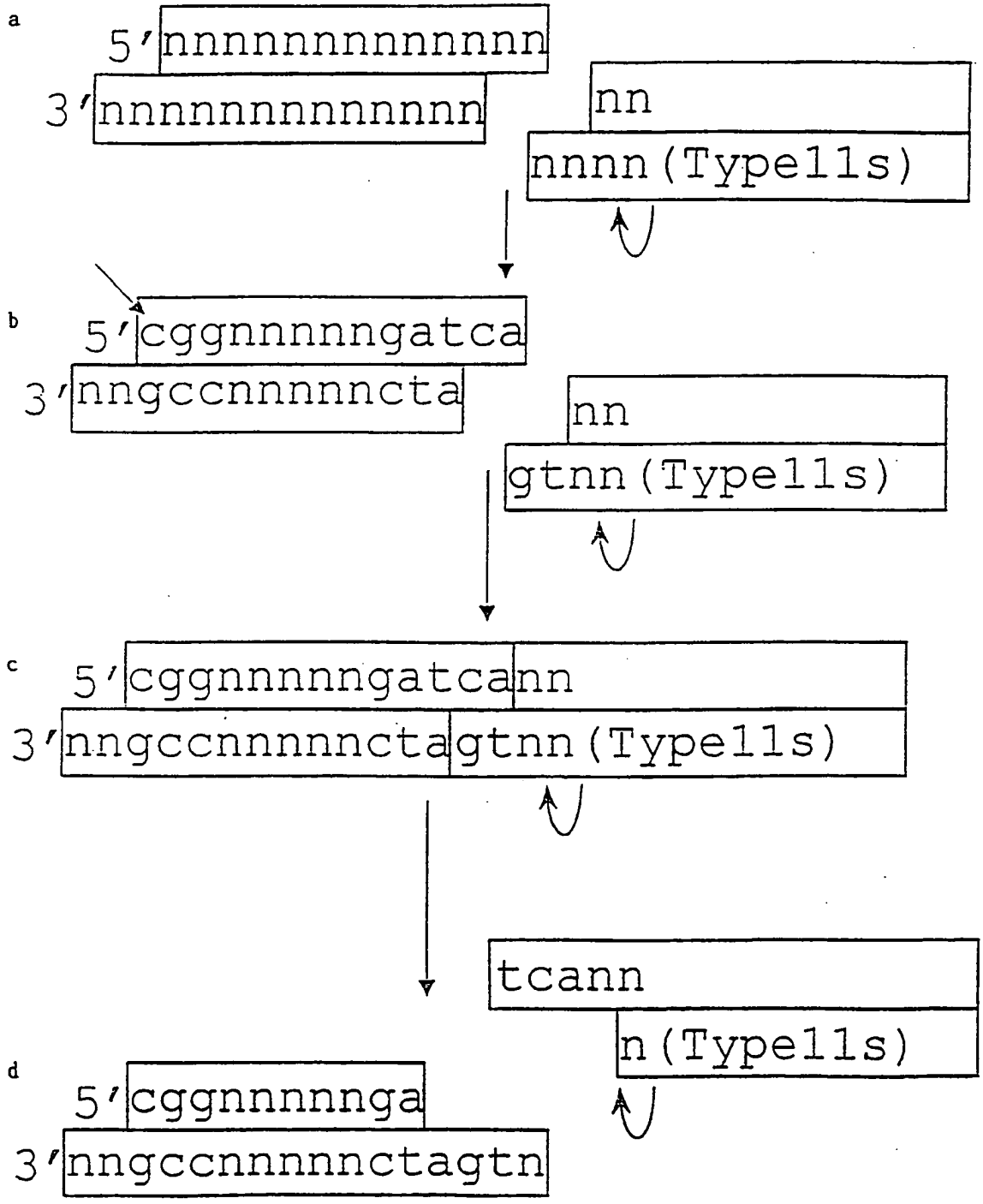

a
5′nnnnnnnnnnnnn
3′nnnnnnnnnnnnn
nn
nnnn(Type11s)
b
5′cggnnnnngatca
3′nngccnnnnnncta
nn
gtnn(Type11s)
c
5′cggnnnnngatcann
3′nngccnnnnnnctagtnn(Type11s)
tcann
n(Type11s)
d
5′cggnnnnnga
3′nngccnnnnnnctagtn

Figure 14

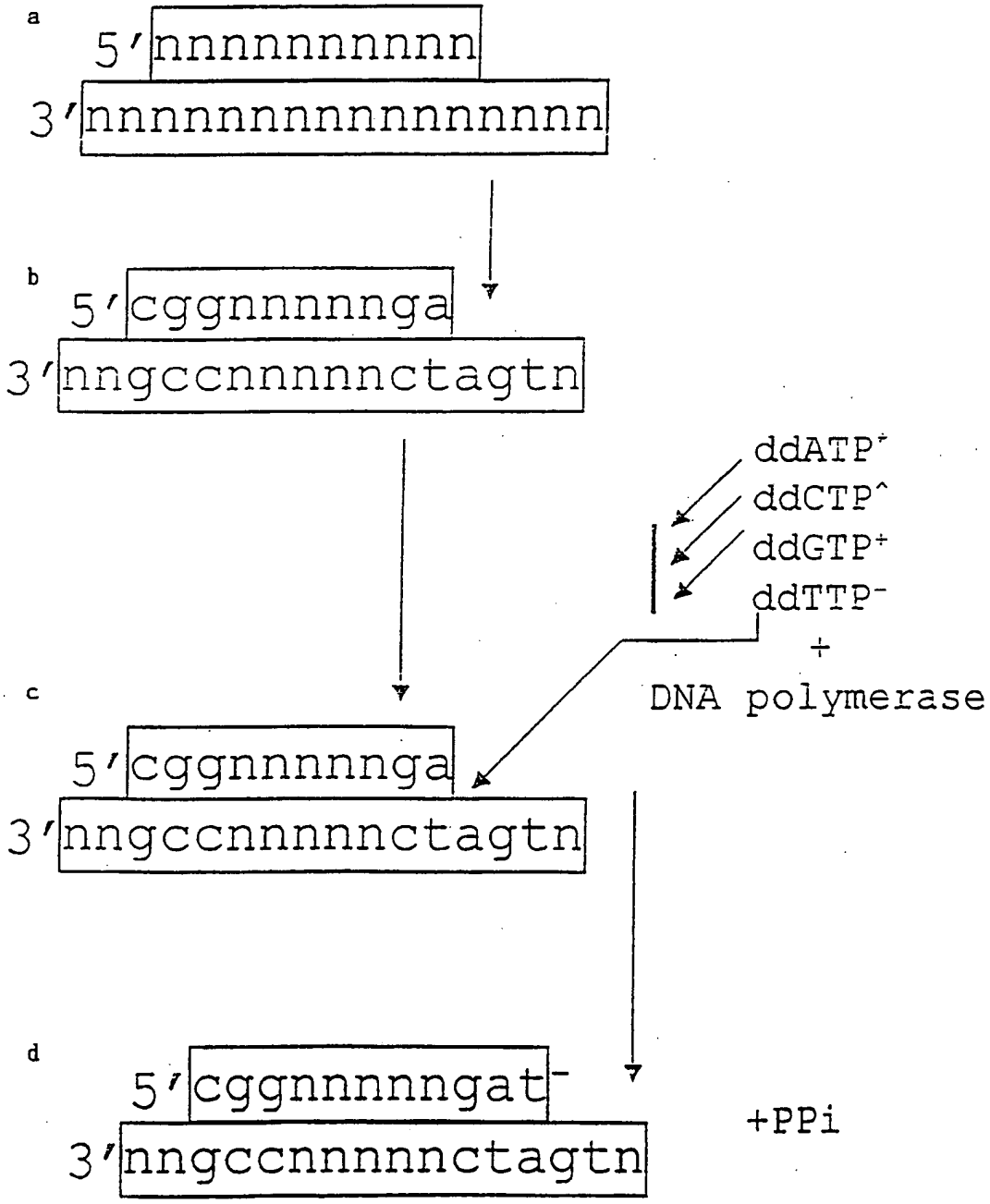

a
5′nnnnnnnnnn
3′nnnnnnnnnnnnnnnn
b
5′cggnnnnnga
3′nngccnnnnnctagtn
ddATP*
ddCTP^
ddGTP+
ddTTP-
+
DNA polymerase
c
5′cggnnnnnga
3′nngccnnnnnctagtn
d
5′cggnnnnngat-
3′nngccnnnnnctagtn
+PPi

Figure 15

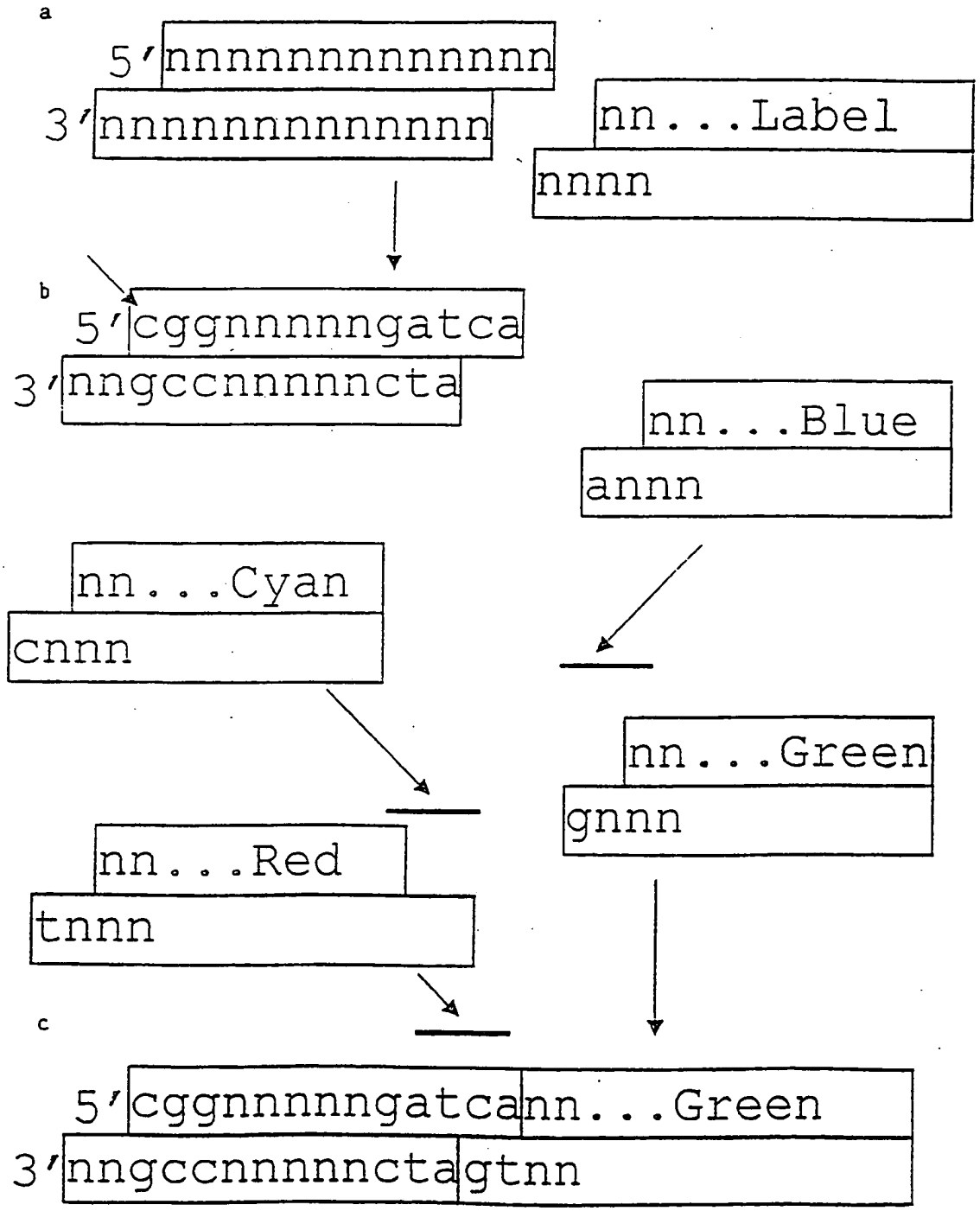

a
5′nnnnnnnnnnnnn
3′nnnnnnnnnnnnn
nn...Label
nnnn
b
5′cggnnnnngatca
3′nngccnnnnnncta
nn...Blue
annn
nn...Cyan
cnnn
nn...Green
gnnn
nn...Red
tnnn
c
5′cggnnnnngatcann...Green
3′nngccnnnnnnctagtnn

Figure 16

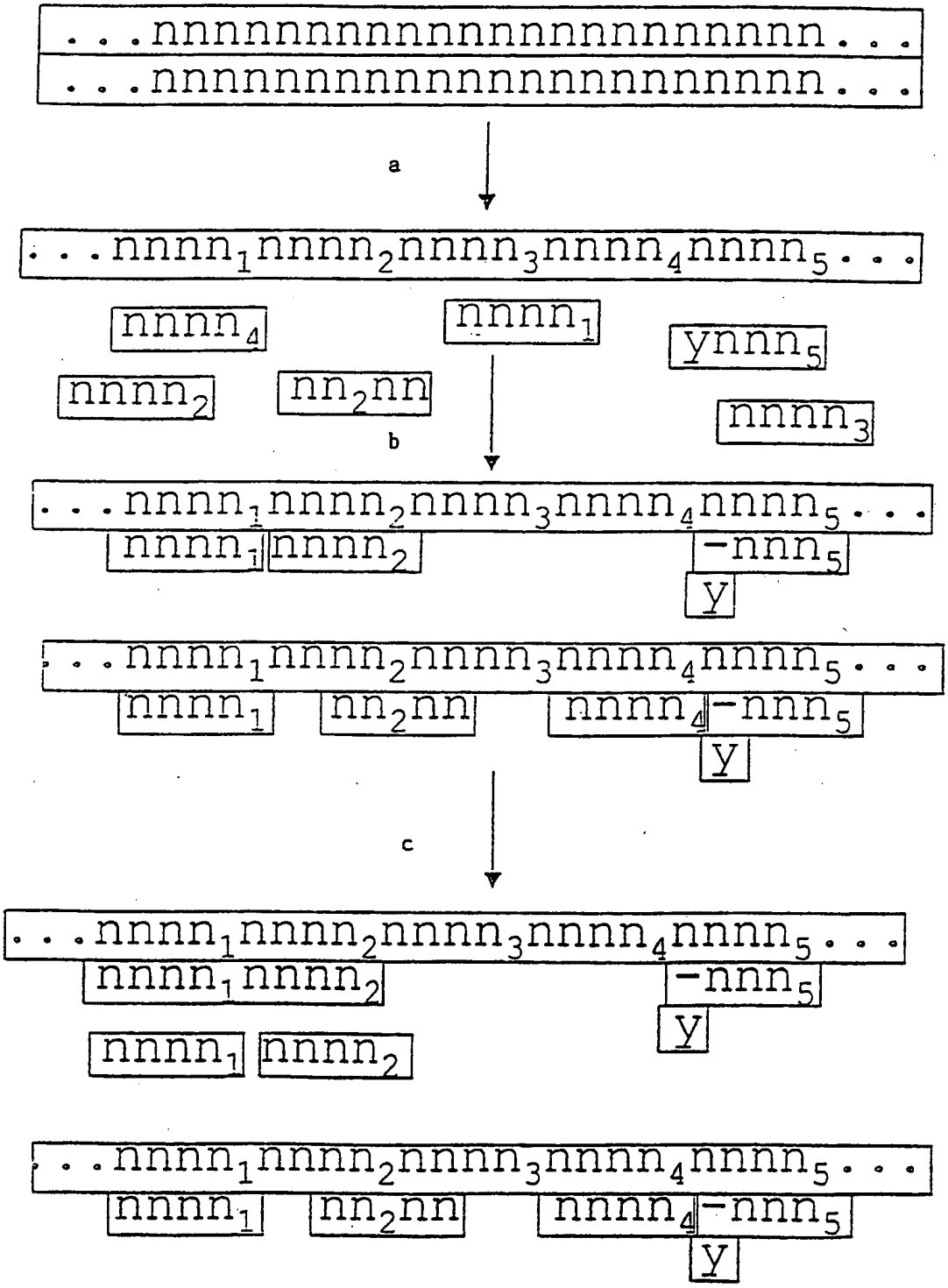

...nnnnnnnnnnnnnnnnnnnnn...
...nnnnnnnnnnnnnnnnnnnnn...
a
...nnnn1nnnn2nnnn3nnnn4nnnn5...
nnnn4
nnnn1
ynnn5
nnnn2
nn2nn
b
nnnn3
...nnnn1nnnn2nnnn3nnnn4nnnn5...
nnnn1 nnnn2
-nnn5
y
...nnnn1nnnn2nnnn3nnnn4nnnn5...
nnnn1
nn2nn
nnnn4 -nnn5
y
c
...nnnn1nnnn2nnnn3nnnn4nnnn5...
nnnn1nnnn2
-nnn5
y
nnnn1
nnnn2
...nnnn1nnnn2nnnn3nnnn4nnnn5...
nnnn1
nn2nn
nnnn4 -nnn5
y

Figure 17

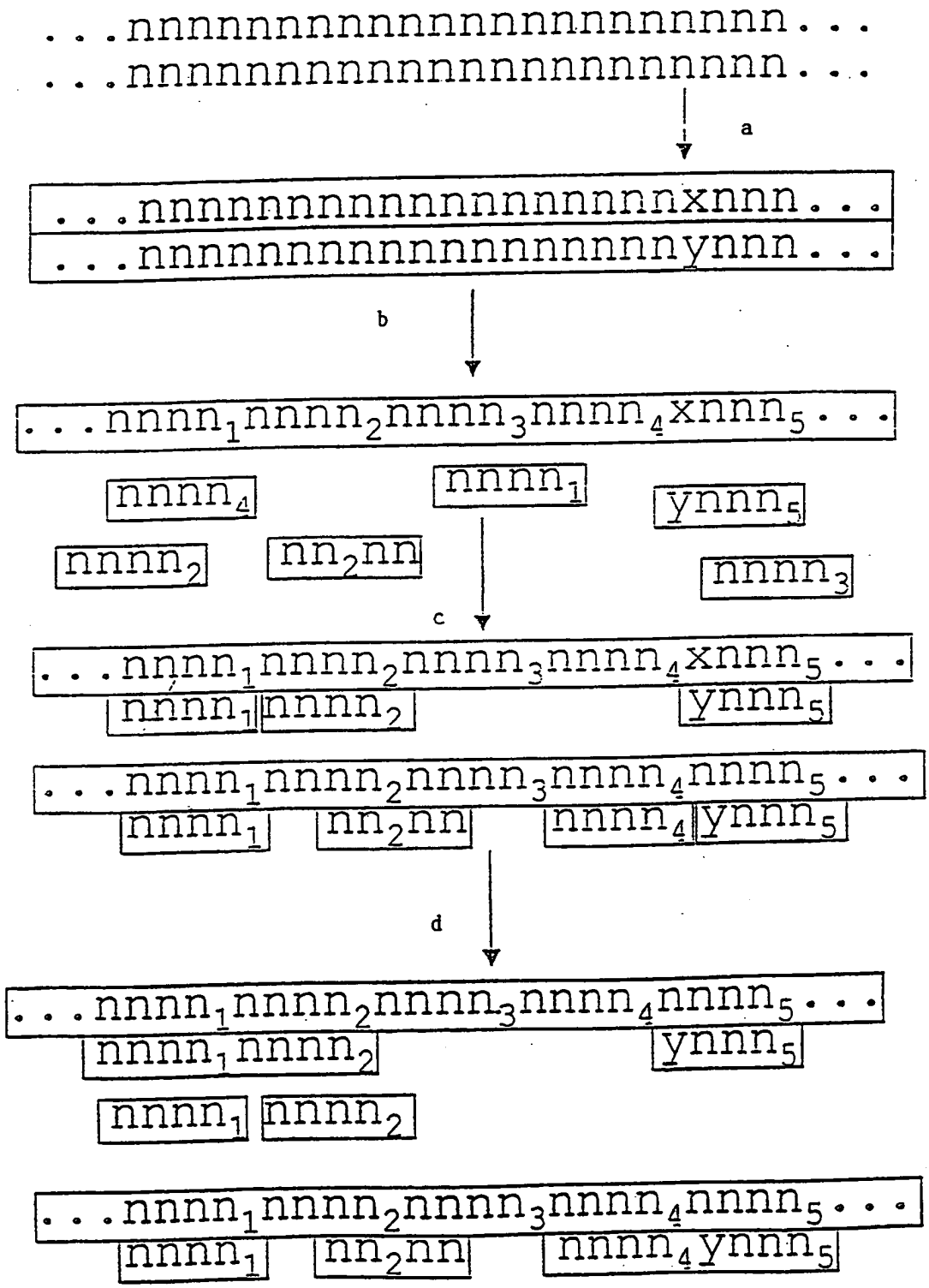

...nnnnnnnnnnnnnnnnnnnnn...
...nnnnnnnnnnnnnnnnnnnnn...
a
...nnnnnnnnnnnnnnnnnnnxnnn...
...nnnnnnnnnnnnnnnnnnnynnn...
b
...nnnn1nnnn2nnnn3nnnn4xnnn5...
nnnn4
nnnn1
ynnn5
nnnn2
nn2nn
nnnn3
c
...nnnn1nnnn2nnnn3nnnn4xnnn5...
nnnn1 nnnn2
ynnn5
...nnnn1nnnn2nnnn3nnnn4nnnn5...
nnnn1
nn2nn
nnnn4 ynnn5
d
...nnnn1nnnn2nnnn3nnnn4nnnn5...
nnnn1nnnn2
ynnn5
nnnn1
nnnn2
...nnnn1nnnn2nnnn3nnnn4nnnn5...
nnnn1
nn2nn
nnnn4ynnn5

Figure 18

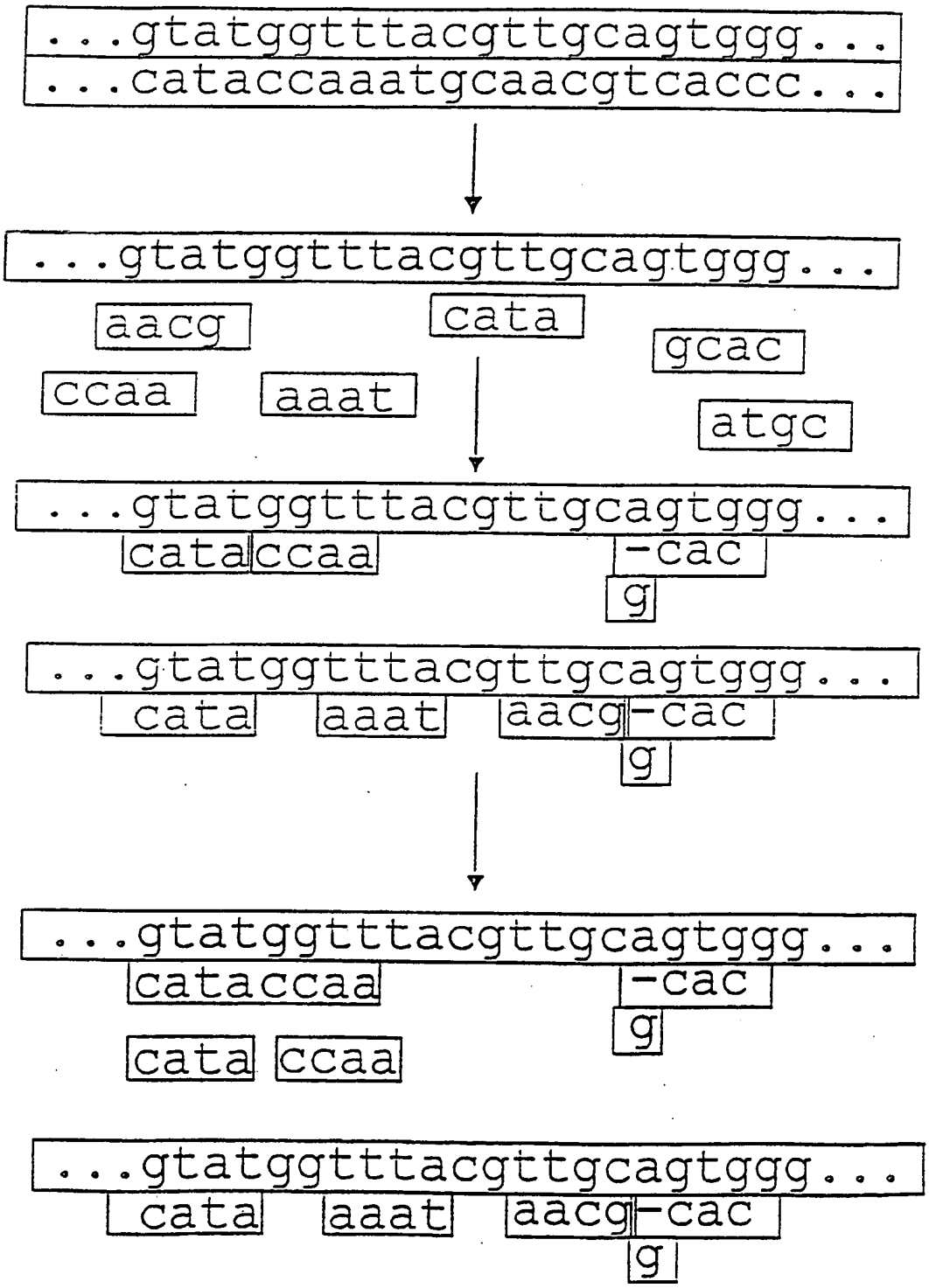

...gtatggtttacgttgcagtggg...
...cataccaaatgcaacgtcaccc...
...gtatggtttacgttgcagtggg...
aacg
cata
gcac
ccaa
aaat
atgc
...gtatggtttacgttgcagtggg...
cata ccaa
-cac
g
...gtatggtttacgttgcagtggg...
cata
aaat
aacg -cac
g
...gtatggtttacgttgcagtggg...
cataccaa
-cac
g
cata
ccaa
...gtatggtttacgttgcagtggg...
cata
aaat
aacg -cac
g

Figure 19

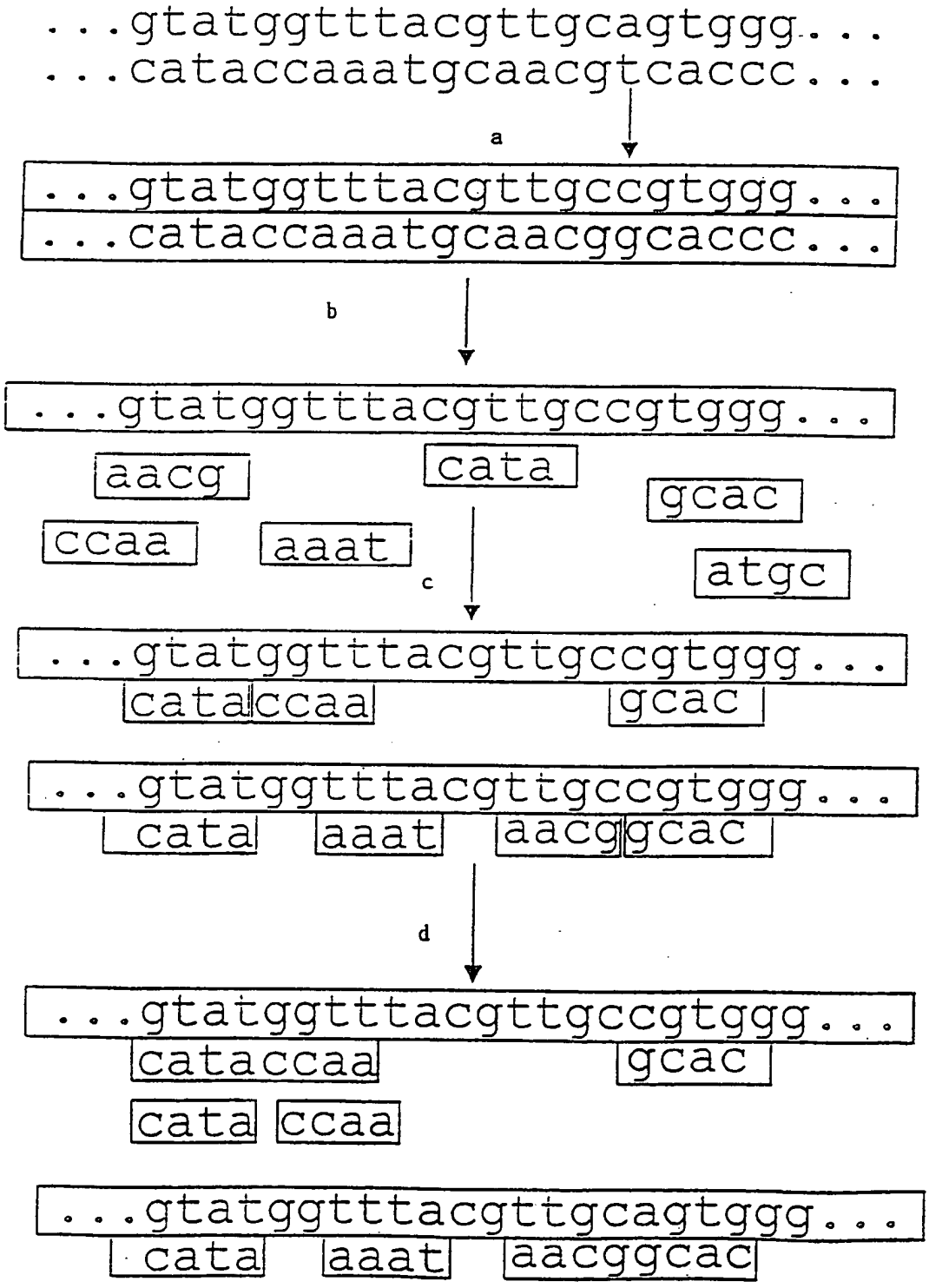
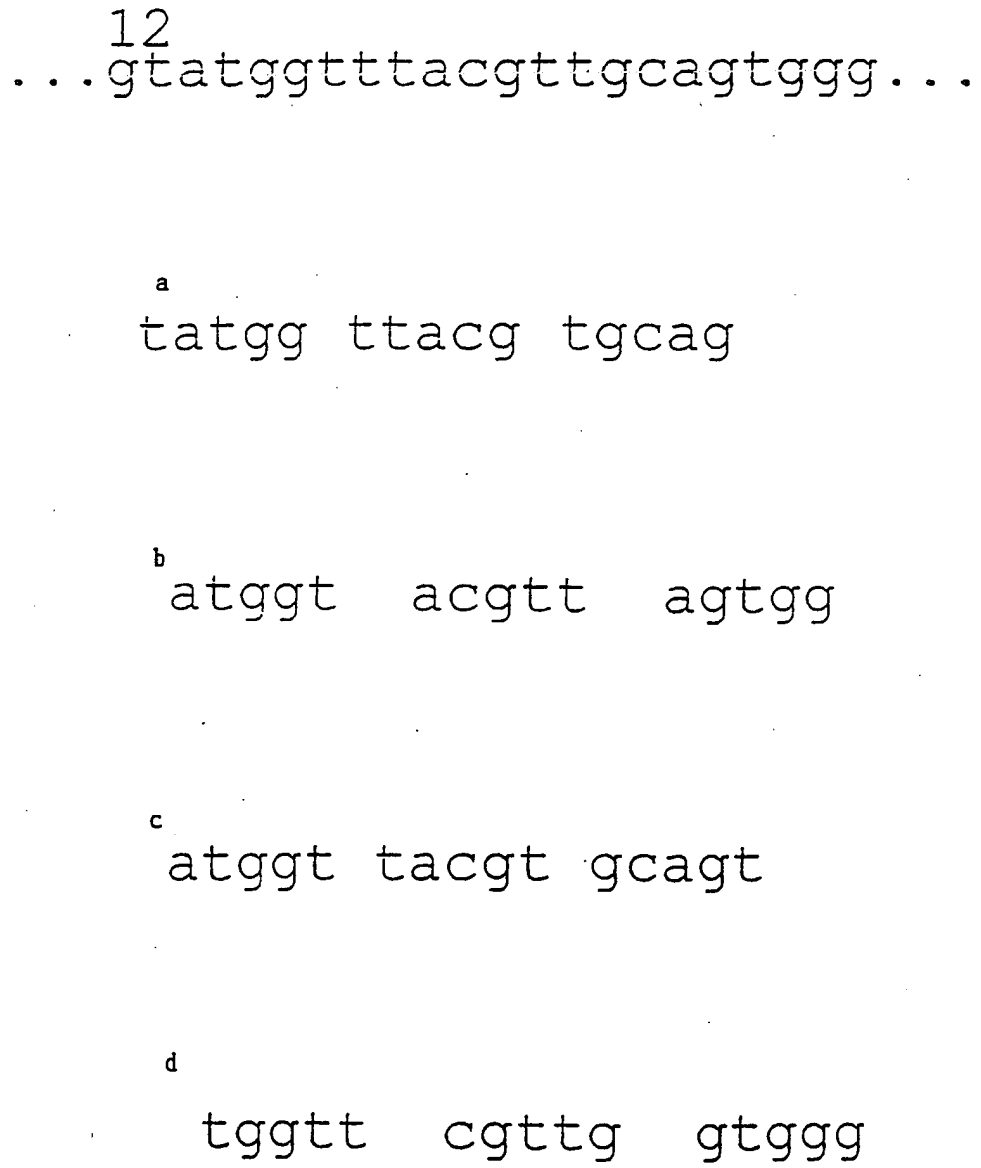

...gtatggtttacgttgcagtggg...
...cataccaaatgcaacgtcaccc...
a
...gtatggtttacgttgccgtggg...
...cataccaaatgcaacggcaccc...
b
...gtatggtttacgttgccgtggg...
aacg
cata
gcac
ccaa
aaat
atgc
c
...gtatggtttacgttgccgtggg...
cata ccaa
gcac
...gtatggtttacgttgccgtggg...
cata
aaat
aacg gcac
d
...gtatggtttacgttgccgtggg...
cataccaa
gcac
cata
ccaa
...gtatggtttacgttgcagtggg...
cata
aaat
aacggcac

Figure 20


12
...gtatggtttacgttgcagtggg...
a
tatgg ttacg tgcag
b
atggt acgtt agtgg
c
atggt tacgt gcagt
d
tggtt cgttg gtggg

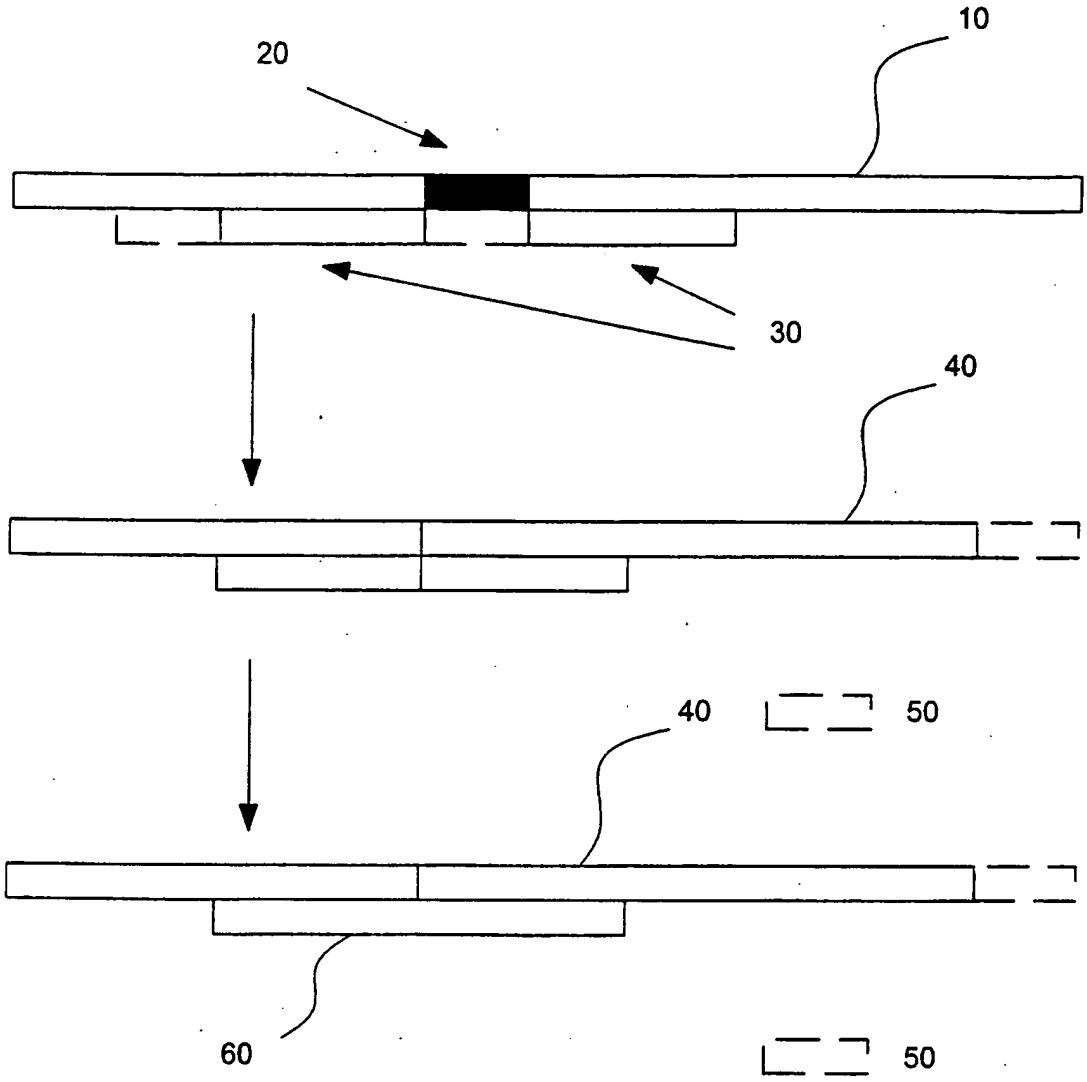


Figure 21

Figure 22

Fragment	1 base end addition	2 base end addition	1 base internal addition @ 2-3	2 base internal addition @ 2-3
gtatgg	agtatgg	aagtatgg	gtaatgg	gtaaatgg
		acgtatgg		gtacatgg
		aggtatgg		gtagatgg
		atgtatgg		gtatatgg
	cgtatgg	cagtatgg	gtcatgg	gtcaatgg
		ccgtatgg		gtccatgg
		cggtatgg		gtcgatgg
		ctgtatgg		gtctatgg
	ggtatgg	gagtatgg	gtgatgg	gtgaatgg
		gcgtatgg		gtgcatgg
		gggtatgg		gtggatgg
		gtgtatgg		gtgtatgg
	tgtatgg	tagtatgg	gttatgg	gttaatgg
		tcgtatgg		gttcatgg
		tggtatgg		gttgatgg
		ttgtatgg		gtttatgg
tttacg	atttacg	aatttacg	ttatacg	ttaatacg
		actttacg		ttactacg
		agtttacg		ttagtacg
		attttacg		ttattacg
	ctttacg	catttacg	ttctacg	ttcatacg
		cctttacg		ttcctacg
		cgtttacg		ttcgtacg
		ctttacg		ttcttacg
	gtttacg	gatttacg	ttgtacg	ttgatacg
		gctttacg		ttgctacg
		ggtttacg		ttggtacg
		gttttacg		ttgttacg
	ttttacg	tatttacg	ttttacg	tttatacg
		tctttacg		tttctacg
		tgtttacg		tttgtacg
		ttttacg		ttttacg

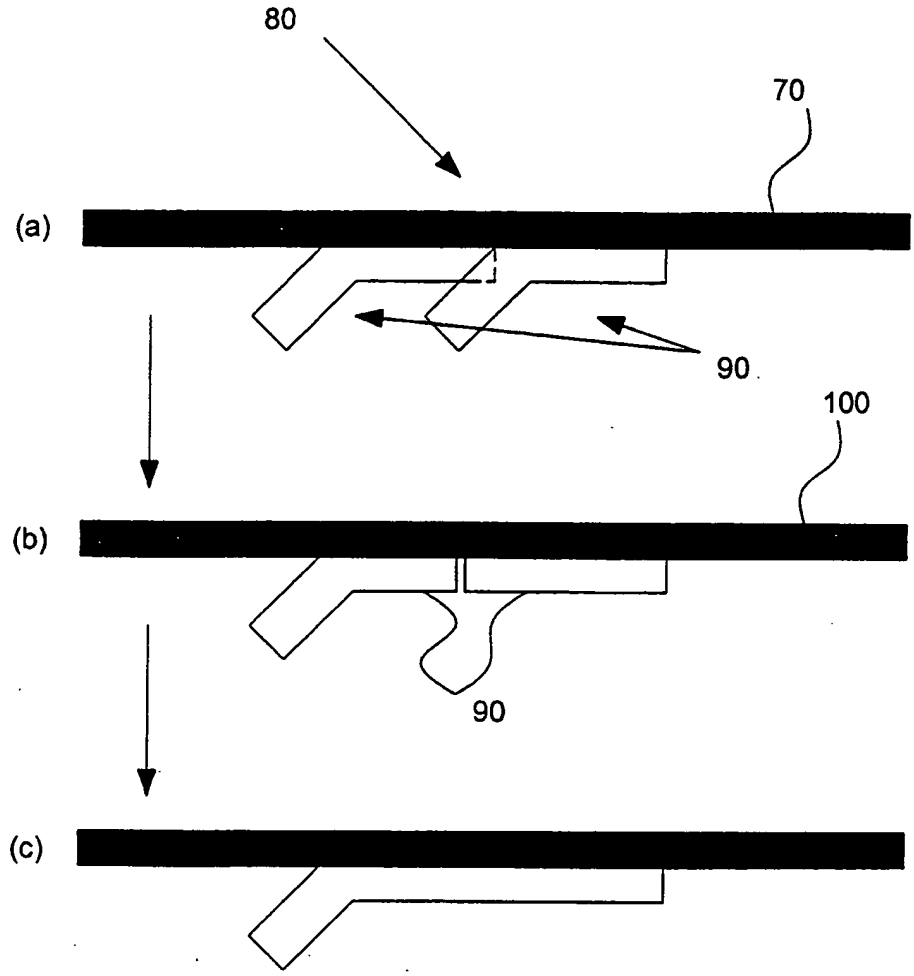


Figure 23

	Target Sequences
1	ctatggaggtatggtttacgc
2	ctatggagatatggtttacgc
3	ctatggaggcatggtttacgc
4	ctatggaggtaaggtttacgc
5	ctatggagtatggtttacgc
6	ctatggagatggtttacgc
7	ctatggaggtaatggtttacgc
8	ctatggaggtacatggtttacgc

Sequence 1 fragment libraries
(Non-mutant target)

4 base fragment library

6 base library fragments	End base substitution fragments	Labelled base fragments	c t a t g g a g g t a t g g t a t g g a g g t a t g g t a t g g a g g t a t g g t t t g g a g g t a t g g t t t
ggaggt	agaggt	agaggt* agaggta* agaggtat*	
	cgaggt	cgaggt* cgaggta* cgaggtat*	
	tgaggt	tgaggt* tgaggta* tgaggtat*	
gaggta	aaggta	aaggta* aaggtat* aaggtatg*	
	caggta	caggta* caggtat* caggtatg*	
	taggta	taggta* taggtat* taggtatg*	

Figure 24(i)

4 base fragment library (continu

6 base library fragments	End base substitution fragments	Labelled base fragments	c t a t g g a g g t a t g g t a t g g a g g t a t g g t a t g g a g g t a t g g t t t g g a g g t a t g g t t t
atggtat	cggtat	cggtat*	
		cggtatg*	
		cggtatgg*	
	gggtat	gggtat*	
		gggtatg*	
		gggtatgg*	
	tggtat	tggtat*	
		tggtatg*	
		tggtatgg*	
ggtatg	agtatg	agtatg*	5
		agtatgg*	5
		agtatggt*	5
	cgtatg	cgtatg*	
		cgtatgg*	
		cgtatggt*	
	tgtatg	tgtatg*	
		tgtatgg*	
		tgtatggt*	
gtatgg	atatgg	atatgg*	2
		atatggt*	2
		atatggtt*	2
	ctatgg	ctatgg*	
		ctatggt*	
		ctatggtt*	
	ttatgg	ttatgg*	
		ttatggt*	
tatggt	aatggt	aatggt*	7
		aatggtt*	7
		aatggttt*	7
	catggt	catggt*	3
		catggtt*	3
		catggttt*	3
	gatggt	gatggt*	6

Figure 24(ii)

		gatggtt*	6
		gatggttt*	6

4 base fragment library (continu

6 base library fragments	End base substitution fragments	Labelled base fragments	c t a t g g a g g t a t g g t a t g g a g g t a t g g t a t g g a g g t a t g g t t t g g a g g t a t g g t t t
atggtt	ctggtt	ctggtt*	
		ctggttt*	
		ctggttta*	
	gtggtt	gtggtt*	
		gtggttt*	
		gtggttta*	
	ttggtt	ttggtt*	
		ttggttt*	
		ttggttta*	
tggttt	aggttt	aggttt*	4
		aggttta*	4
		aggtttac*	4
	cggttt	cggttt*	
		cggttta*	
		cggtttac*	
	gggttt	gggttt*	
		gggttta*	
		gggtttac*	
ggttta	agttta	agttta*	
		agtttac*	
		agtttacg*	
	cgttta	cgttta*	
		cgtttac*	
		cgtttacg*	
	tgttta	tgttta*	
		tgtttac*	
		tgtttacg*	
gtttac	atttac	atttac*	
		atttacg*	
		atttacgc*	
	ctttac	ctttac*	
		ctttacg*	

Figure 24(iii)

	ttttac	ctttacgc* ttttac* ttttacg* ttttacgc*		

			4 base fragment library	4 base fragment library
			Odd positions c a g a g a g t t g g t t g a g a g a g t t g g t t g t	Even positions t t g g t t g a g a g a g t t g g t t g t g a g a g t t
ggaggt	agaggt cgaggt tgaggt	agaggt* agaggta* agaggtat* cgaggt* cgaggta* cgaggtat* tgaggt* tgaggta* tgaggtat*		
gaggta	aaggta caggta taggta	aaggta* aaggtat* aaggtatg* caggta* caggtat* caggtatg* taggta* taggtat* taggtatg*		
aggtat	cggtat gggtat tggtat	cggtat* cggtatg* cggtatgg* gggtat* gggtatg* gggtatgg* tggtat* tggtatg* tggtatgg*		
ggtatg	agtatg	agtatg*	5	

Figure 24(iv)

	cgtatg	agtatgg* agtatggt* cgtatg* cgtatgg* cgtatggt*	5 5	

Figure 24(v)

			4 base fragment library	4 base fragment library
			Odd positions (cont.	Even positions (cont.)
			c a g a g a g t t g g t t g a g a g a g t t g g t t g t	t t g g t t g a g a g a g t t g g t t g t g a g a g t t
	tgtatg	tgtatg* tgtatgg* tgtatggt*		
gtatgg	atatgg	atatgg* atatggt* atatggtt*	2 2 2	
	ctatgg	ctatgg* ctatggt* ctatggtt*		
	ttatgg	ttatgg* ttatggt* ttatggtt*		
tatggt	aatggt	aatggt* aatggtt* aatggttt*	7 7 7	
	catggt	catggt* catggtt* catggttt*		3 3 3
	gatggt	gatggt* gatggtt* gatggttt*		6 6 6
atggtt	ctggtt	ctggtt* ctggttt* ctggttta*		
	gtggtt	gtggtt* gtggttt* gtggttta*		
	ttggtt	ttggtt* ttggttt* ttggttta*		
tggttt	aggttt	aggttt* aggttta* aggtttac*		4 4 4

Figure 24(vi)

			4 base fragment library	4 base fragment library
			Odd positions (cont.	Even positions (cont.)
			c a g a g a g t t g g t t g a g a g a g t t g g t t g t	t t g g t t g a g a g a g t t g g t t g t g a g a g t t
	cggttt gggttt	cggttt* cggttta* cggtttac* gggttt* gggttta* gggtttac*		
ggttta	agttta cgttta tgttta	agttta* agtttac* agtttacg* cgttta* cgtttac* cgtttacg* tgttta* tgtttac* tgtttacg*		
gtttac	atttac ctttac ttttac	atttac* atttacg* atttacgc* ctttac* ctttacg* ctttacgc* ttttac* ttttacg* ttttacgc*		

Figure 24(vii)

```
5` | g g g a t c c t t g t c g a a t a a a g t c g a g g t g c t a g t t t c a t a a g c a a a | 3`
     |c a g c t t a t t t c a g c|t c c a c g a t c a a a g t a t 5`|
     |<----Oligonucleotide B---->|<--------Oligonucleotide U---------->|
                  ligation point
```

Figure 25

SEQUENCE LISTING

<110> Clatterbridge Cancer Research Trust

<120> GENETIC ANALYSIS

<130> M99/0734/PCT

<140>
<141>

<150> GB9927520.8
<151> 1999-11-23

<150> GB9906833.0
<151> 1999-03-24

<160> 259

<170> PatentIn Ver. 2.1

<210> 1
<211> 11
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 1
attgcgcatt g 11

<210> 2
<211> 10
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 2
attgcgattg 10

<210> 3

<211> 10
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 3
attgccattg 10

<210> 4
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 4
gggatcttgt cgaataaagt cgaggtgcta gtttcataag caaa 44

<210> 5
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 5
gggatcttgt cgaataaagt ctaggtgcta gtttcataag caaa 44

<210> 6
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 6
gggatcttgt cgaataaagt ccaggtgcta gtttcataag caaa 44

<210> 7
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 7
gggatcttgt cgaataaagt caggtgctag tttcataagc aaa 43

<210> 8
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 8
gggatcttgt cgaataaagt cgaaggtgct agtttcataa gcaaa 45

<210> 9
<211> 45
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 9
gggatcttgt cgaataaagt cgatggtgct agtttcataa gcaaa 45

<210> 10
<211> 13
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 10
agactttatt cga 13

<210> 11
<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 11
cgactttatt cgac 14

<210> 12
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 12
ggactttatt cgaca 15

<210> 13
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 13

tgactttatt cgaca 15

<210> 14
<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 14
agactttatt cgac 14

<210> 15
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 15
cgactttatt cga 13

<210> 16
<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 16
ggactttatt cgac 14

<210> 17
<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 17
tgactttatt cgac 14

<210> 18
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 18
agactttatt cgaca 15

<210> 19
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 19
aactttattc gacaag 16

<210> 20
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 20
cactttattc gacaag 16

<210> 21
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 21
gactttattc gacaag 16

<210> 22
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 22
tactttattc gacaag 16

<210> 23
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 23
acgactttat tcgacaag 18

<210> 24
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 24
ccgactttat tcgacaaga 19

<210> 25
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 25
gcgactttat tcgacaagat 20

<210> 26
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 26
tcgactttat tcgacaagat c 21

<210> 27
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 27
atcgacttta ttc 13

<210> 28
<211> 15
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 28
ctcgacttta ttcga 15

<210> 29
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 29
gtcgacttta ttcga 15

<210> 30
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 30
ttcgacttta ttcga 15

<210> 31
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 31

tatgaaacta gcacct 16

<210> 32
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 32
atgaaactag cacct 15

<210> 33
<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 33
tgaaactagc acct 14

<210> 34
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 34
gcttatgaaa ctagcacc 18

<210> 35
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 35
cttatgaaac tagcacc 17

<210> 36
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 36
ttatgaaact agcacc 16

<210> 37
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 37
ttccagttgc tttatctgtt ca 22

<210> 38
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 38
aagagcaatc agtgaggaat caga 24

<210> 39

<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Adapter

<400> 39
ccagtcgcag gtctcaagct cgacagctgg agnn 34

<210> 40
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 40
acggtagcta aagctgagtc cccgttgtag 30

<210> 41
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 41
nnnnnnnnnn nnnnnn 16

<210> 42
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 42

nngtcatggt tccagnnn 18

<210> 43
<211> 10
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 43
nnatggttnn 10

<210> 44
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 44
cggnnnnnga tcann 15

<210> 45
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 45
nntgatcnnn nnccgnn 17

<210> 46
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 46
gtatggttta cgttgcagtg gg 22

<210> 47
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 47
gtatggttta cgttgccgtg gg 22

<210> 48
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 48
ctatggaggt atggtttacg c 21

<210> 49
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 49
ctatggagat atggtttacg c 21

<210> 50
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 50
ctatggaggc atggtttacg c 21

<210> 51
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 51
ctatggaggt aaggtttacg c 21

<210> 52
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 52
ctatggagta tggtttacgc 20

<210> 53
<211> 19
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 53
ctatggagat ggtttacgc 19

<210> 54
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 54
ctatggaggt aatggtttac gc 22

<210> 55
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 55
ctatggaggt acatggttta cgc 23

<210> 56
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 56
gggatcttgt cgaataaagt cgaggtgcta gtttcataag caaa 44

<210> 57
<211> 14
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 57
cgactttatt cgac 14

<210> 58
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Arbitrary example sequence

<400> 58
tatgaaacta gcacct 16

<210> 59
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 59
tctcaacagc ggtaagatcc 20

<210> 60
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 60
caacagcggt aagatccttg 20

<210> 61
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 61
acgctcaccg gcaccagatt 20

<210> 62
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 62
tcaccggcac cagatttatc 20

<210> 63
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 63
cctgtagcta tggcaacaac 20

<210> 64
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 64
atgcctgtag ctatggcaac 20

<210> 65
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 65
tgcctgtagc tatggcaaca 20

<210> 66
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 66
agtatttggt atctgcgctc 20

<210> 67
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 67
tatttggtat ctgcgctctg 20

<210> 68
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 68

ttggtatctg cgctctgctg 20

<210> 69
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 69
gtatctgcgc tctgctgaag 20

<210> 70
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 70
ggtaatacgg ttatccacag 20

<210> 71
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 71
caacagcggt aagatccttg 20

<210> 72
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 72
ctcgcggtat aattgcagca 20

<210> 73
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 73
tctcgcggta taattgcagc 20

<210> 74
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 74
gtctcgcggt ataattgcag 20

<210> 75
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 75
tgctgcaatt ataccgcgag 20

<210> 76
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: PCR primer

<400> 76
ctgcaattat accgcgagac 20

<210> 77
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 77
aattctggag aacattgccg acaaggatcc 30

<210> 78
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 78
aattctggag accattgccg acaaggatcc 30

<210> 79
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 79
aattctggag agcattgccg acaaggatcc 30

<210> 80

<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 80
aattctggag atcattgccg acaaggatcc 30

<210> 81
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 81
aattctggag cacattgccg acaaggatcc 30

<210> 82
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 82
aattctggag cccattgccg acaaggatcc 30

<210> 83
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 83
aattctggag cgcattgccg acaaggatcc 30

<210> 84
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 84
aattctggag ctcattgccg acaaggatcc 30

<210> 85
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 85
aattctggag gacattgccg acaaggatcc 30

<210> 86
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 86
aattctggag gccattgccg acaaggatcc 30

<210> 87
<211> 30
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 87
aattctggag ggcattgccg acaaggatcc 30

<210> 88
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 88
aattctggag gtcattgccg acaaggatcc 30

<210> 89
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 89
aattctggag tacattgccg acaaggatcc 30

<210> 90
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 90

aattctggag tccattgccg acaaggatcc 30

<210> 91
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 91
aattctggag tgcattgccg acaaggatcc 30

<210> 92
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 92
aattctggag ttcattgccg acaaggatcc 30

<210> 93
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 93
agctggatcc ttgtcggcaa tgttctccag 30

<210> 94
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 94
agctggatcc ttgtcggcaa tggtctccag 30

<210> 95
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 95
agctggatcc ttgtcggcaa tgctctccag 30

<210> 96
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 96
agctggatcc ttgtcggcaa tgatctccag 30

<210> 97
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 97
agctggatcc ttgtcggcaa tgtgctccag 30

<210> 98
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 98
agctggatcc ttgtcggcaa tgggctccag 30

<210> 99
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 99
agctggatcc ttgtcggcaa tgcgctccag 30

<210> 100
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 100
agctggatcc ttgtcggcaa tgagctccag 30

<210> 101
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 101
agctggatcc ttgtcggcaa tgtcctccag 30

<210> 102
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 102
agctggatcc ttgtcggcaa tggcctccag 30

<210> 103
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 103
agctggatcc ttgtcggcaa tgccctccag 30

<210> 104
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 104
agctggatcc ttgtcggcaa tgacctccag 30

<210> 105
<211> 30
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 105
agctggatcc ttgtcggcaa tgtactccag 30

<210> 106
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 106
agctggatcc ttgtcggcaa tggactccag 30

<210> 107
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 107
agctggatcc ttgtcggcaa tgcactccag 30

<210> 108
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 108

agctggatcc ttgtcggcaa tgaactccag 30

<210> 109
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Replacement polylinker

<400> 109
agctggatcc ttgtcggcaa tgnnctccag 30

<210> 110
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 110
aggcacccca ggctttac 18

<210> 111
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 111
ccgcacagat gcgtaagg 18

<210> 112
<211> 27
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 112
aattcctgga gnnnnnnnnn nnnnnaa 27

<210> 113
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 113
aattcctgga gnnnnnnnnn nnnnnac 27

<210> 114
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 114
aattcctgga gnnnnnnnnn nnnnnag 27

<210> 115
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 115
aattcctgga gnnnnnnnnn nnnnnat 27

<210> 116
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 116
aattcctgga gnnnnnnnnn nnnnnca 27

<210> 117
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 117
aattcctgga gnnnnnnnnn nnnnncc 27

<210> 118
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 118
aattcctgga gnnnnnnnnn nnnnncg 27

<210> 119
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 119
aattcctgga gnnnnnnnnn nnnnnct 27

<210> 120
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 120
aattcctgga gnnnnnnnnn nnnnnga 27

<210> 121
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 121
aattcctgga gnnnnnnnnn nnnnngc 27

<210> 122
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 122
aattcctgga gnnnnnnnnn nnnnngg 27

<210> 123
<211> 27
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 123
aattcctgga gnnnnnnnnn nnnnngt 27

<210> 124
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 124
aattcctgga gnnnnnnnnn nnnnnta 27

<210> 125
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 125
aattcctgga gnnnnnnnnn nnnnntc 27

<210> 126
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 126

aattcctgga gnnnnnnnnn nnnnntg 27

<210> 127
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 127
aattcctgga gnnnnnnnnn nnnnntt 27

<210> 128
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 128
nnnnnnnnnn nnnnctccag g 21

<210> 129
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 129
aattccctgg agnnnnnnnn nnnnnaa 27

<210> 130
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 130
aattccctgg agnnnnnnnn nnnnnac 27

<210> 131
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 131
aattccctgg agnnnnnnnn nnnnnag 27

<210> 132
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 132
aattccctgg agnnnnnnnn nnnnnat 27

<210> 133
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 133
aattccctgg agnnnnnnnn nnnnnca 27

<210> 134
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 134
aattccctgg agnnnnnnnn nnnnncc 27

<210> 135
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 135
aattccctgg agnnnnnnnn nnnnncg 27

<210> 136
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 136
aattccctgg agnnnnnnnn nnnnnct 27

<210> 137
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 137
aattccctgg agnnnnnnnn nnnnnga 27

<210> 138
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 138
aattccctgg agnnnnnnnn nnnnngc 27

<210> 139
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 139
aattccctgg agnnnnnnnn nnnnngg 27

<210> 140
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 140
aattccctgg agnnnnnnnn nnnnngt 27

<210> 141
<211> 27
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 141
aattccctgg agnnnnnnnn nnnnnta 27

<210> 142
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 142
aattccctgg agnnnnnnnn nnnnntc 27

<210> 143
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 143
aattccctgg agnnnnnnnn nnnnntg 27

<210> 144
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 144

aattccctgg agnnnnnnnn nnnnntt 27

<210> 145
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 145
nnnnnnnnnn nnnctccagg g 21

<210> 146
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 146
aattctggag nnnnnnnnnn nnnnaaa 27

<210> 147
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 147
aattctggag nnnnnnnnnn nnnnaac 27

<210> 148
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 148
aattctggag nnnnnnnnnn nnnnaag 27

<210> 149
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 149
aattctggag nnnnnnnnnn nnnnaat 27

<210> 150
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 150
aattctggag nnnnnnnnnn nnnnaca 27

<210> 151
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 151
aattctggag nnnnnnnnnn nnnnacc 27

<210> 152
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 152
aattctggag nnnnnnnnnn nnnnacg 27

<210> 153
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 153
aattctggag nnnnnnnnnn nnnnact 27

<210> 154
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 154
aattctggag nnnnnnnnnn nnnnaga 27

<210> 155
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 155
aattctggag nnnnnnnnnn nnnnagc 27

<210> 156
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 156
aattctggag nnnnnnnnnn nnnnagg 27

<210> 157
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 157
aattctggag nnnnnnnnnn nnnnagt 27

<210> 158
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 158
aattctggag nnnnnnnnnn nnnnata 27

<210> 159
<211> 27
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 159
aattctggag nnnnnnnnnn nnnnatc 27

<210> 160
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 160
aattctggag nnnnnnnnnn nnnnatg 27

<210> 161
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 161
aattctggag nnnnnnnnnn nnnnatt 27

<210> 162
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 162

aattctggag nnnnnnnnnn nnnncaa 27

<210> 163
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 163
aattctggag nnnnnnnnnn nnnncac 27

<210> 164
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 164
aattctggag nnnnnnnnnn nnnncag 27

<210> 165
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 165
aattctggag nnnnnnnnnn nnnncat 27

<210> 166
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 166
aattctggag nnnnnnnnnn nnnncca 27

<210> 167
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 167
aattctggag nnnnnnnnnn nnnnccc 27

<210> 168
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 168
aattctggag nnnnnnnnnn nnnnccg 27

<210> 169
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 169
aattctggag nnnnnnnnnn nnnncct 27

<210> 170
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 170
aattctggag nnnnnnnnnn nnnncga 27

<210> 171
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 171
aattctggag nnnnnnnnnn nnnncgc 27

<210> 172
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 172
aattctggag nnnnnnnnnn nnnncgg 27

<210> 173
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 173
aattctggag nnnnnnnnnn nnnncgt 27

<210> 174
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 174
aattctggag nnnnnnnnnn nnnncta 27

<210> 175
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 175
aattctggag nnnnnnnnnn nnnnctc 27

<210> 176
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 176
aattctggag nnnnnnnnnn nnnnctg 27

<210> 177
<211> 27
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 177
aattctggag nnnnnnnnnn nnnnctt 27

<210> 178
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 178
aattctggag nnnnnnnnnn nnnngaa 27

<210> 179
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 179
aattctggag nnnnnnnnnn nnnngac 27

<210> 180
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 180

aattctggag nnnnnnnnnn nnnngag 27

<210> 181
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 181
aattctggag nnnnnnnnnn nnnngat 27

<210> 182
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 182
aattctggag nnnnnnnnnn nnnngca 27

<210> 183
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 183
aattctggag nnnnnnnnnn nnnngcc 27

<210> 184
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 184
aattctggag nnnnnnnnnn nnnngcg 27

<210> 185
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 185
aattctggag nnnnnnnnnn nnnngct 27

<210> 186
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 186
aattctggag nnnnnnnnnn nnnngga 27

<210> 187
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 187
aattctggag nnnnnnnnnn nnnnggc 27

<210> 188
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 188
aattctggag nnnnnnnnnn nnnnggg 27

<210> 189
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 189
aattctggag nnnnnnnnnn nnnnggt 27

<210> 190
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 190
aattctggag nnnnnnnnnn nnnngta 27

<210> 191
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 191
aattctggag nnnnnnnnnn nnnngtc 27

<210> 192
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 192
aattctggag nnnnnnnnnn nnnngtg 27

<210> 193
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 193
aattctggag nnnnnnnnnn nnnngtt 27

<210> 194
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 194
aattctggag nnnnnnnnnn nnnntaa 27

<210> 195
<211> 27
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 195
aattctggag nnnnnnnnnn nnnntac 27

<210> 196
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 196
aattctggag nnnnnnnnnn nnnntag 27

<210> 197
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 197
aattctggag nnnnnnnnnn nnnntat 27

<210> 198
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 198

aattctggag nnnnnnnnnn nnnntca 27

<210> 199
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 199
aattctggag nnnnnnnnnn nnnntcc 27

<210> 200
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 200
aattctggag nnnnnnnnnn nnnntcg 27

<210> 201
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 201
aattctggag nnnnnnnnnn nnnntct 27

<210> 202
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 202
aattctggag nnnnnnnnnn nnnntga 27

<210> 203
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 203
aattctggag nnnnnnnnnn nnnntgc 27

<210> 204
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 204
aattctggag nnnnnnnnnn nnnntgg 27

<210> 205
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 205
aattctggag nnnnnnnnnn nnnntgt 27

<210> 206
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 206
aattctggag nnnnnnnnnn nnnntta 27

<210> 207
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 207
aattctggag nnnnnnnnnn nnnnttc 27

<210> 208
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 208
aattctggag nnnnnnnnnn nnnnttg 27

<210> 209
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 209
aattctggag nnnnnnnnnn nnnnttt 27

<210> 210
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 210
tnnnnnnnnn nnnnnctcca g 21

<210> 211
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 211
gnnnnnnnnn nnnnnctcca g 21

<210> 212
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 212
cnnnnnnnnn nnnnnctcca g 21

<210> 213
<211> 21
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 213
annnnnnnnn nnnnnctcca g 21

<210> 214
<211> 64
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 214
gaggctcagt gatacagtct tccacggccg ttgtaaattg tcgggaagac tgctcctcca 60

gcag 64

<210> 215
<211> 64
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 215
gaggctcagg atacagtctt ctcacggccg ttgtaaattg tcggaagact gctccctcca 60

gcag 64

<210> 216
<211> 64
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement

plasmid sequence

<400> 216
gaggctcaga tacagtcttc gtcacggccg ttgtaaattg tcgaagactg ctccgctcca 60

gcag 64

<210> 217
<211> 64
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 217
gaggctcgat acagtcttca gtcacggccg ttgtaaattg tgaagactgc tcccgctcca 60

gcag 64

<210> 218
<211> 64
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 218
gaggctgata cagtcttcca gtcacggccg ttgtaaattg gaagactgct cctcgctcca 60

gcag 64

<210> 219
<211> 64
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 219
gaggcgatac agtcttctca gtcacggccg ttgtaaattg aagactgctc cgtcgctcca 60

gcag 64

<210> 220
<211> 64
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 220
gagggataca gtcttcctca gtcacggccg ttgtaaatga agactgctcc tgtcgctcca 60

gcag 64

<210> 221
<211> 64
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 221
gaggatacag tcttcgctca gtcacggccg ttgtaaagaa gactgctcct tgtcgctcca 60

gcag 64

<210> 222
<211> 64
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 222

gagatacagt cttcggctca gtcacggccg ttgtaagaag actgctccat tgtcgctcca 60

gcag 64

<210> 223
<211> 64
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 223
ggatacagtc ttcaggctca gtcacggccg ttgtagaaga ctgctccaat tgtcgctcca 60

gcag 64

<210> 224
<211> 64
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 224
gatacagtct tcgaggctca gtcacggccg ttgtgaagac tgctccaaat tgtcgctcca 60

gcag 64

<210> 225
<211> 64
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 225
atacagtctt ctgaggctca gtcacggccg ttggaagact gctcctaaat tgtcgctcca 60

gcag 64

<210> 226
<211> 64
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 226
tacagtcttc gtgaggctca gtcacggccg ttgaagactg ctccgtaaat tgtcgctcca 60

gcag 64

<210> 227
<211> 64
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 227
acagtcttcc gtgaggctca gtcacggccg tgaagactgc tcctgtaaat tgtcgctcca 60

gcag 64

<210> 228
<211> 70
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 228
gatcctgctg gaggagcagt cttcccgaca atttacaacg gccgtggaag actgtatcac 60

tgagcctcac 70

<210> 229
<211> 70
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 229
gatcctgctg gagggagcag tcttccgaca atttacaacg gccgtgagaa gactgtatcc 60

tgagcctcac 70

<210> 230
<211> 70
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 230
gatcctgctg gagcggagca gtcttcgaca atttacaacg gccgtgacga agactgtatc 60

tgagcctcac 70

<210> 231
<211> 70
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement plasmid sequence

<400> 231
gatcctgctg gagcgggagc agtcttcaca atttacaacg gccgtgactg aagactgtat 60

cgagcctcac 70

<210> 232

<211> 70
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement
plasmid sequence

<400> 232
gatcctgctg gagcgaggag cagtcttcca atttacaacg gccgtgactg gaagactgta 60

tcagcctcac 70

<210> 233
<211> 70
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement
plasmid sequence

<400> 233
gatcctgctg gagcgacgga gcagtcttca atttacaacg gccgtgactg agaagactgt 60

atcgcctcac 70

<210> 234
<211> 70
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement
plasmid sequence

<400> 234
gatcctgctg gagcgacagg agcagtcttc atttacaacg gccgtgactg aggaagactg 60

tatccctcac 70

<210> 235
<211> 70
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement
plasmid sequence

<400> 235
gatcctgctg gagcgacaag gagcagtctt ctttacaacg gccgtgactg agcgaagact 60

gtatcctcac 70

<210> 236
<211> 70
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement
plasmid sequence

<400> 236
catcctgctg gagcgacaat ggagcagtct tcttacaacg gccgtgactg agccgaagac 60

tgtatctcac 70

<210> 237
<211> 70
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement
plasmid sequence

<400> 237
gatcctgctg gagcgacaat tggagcagtc ttctacaacg gccgtgactg agcctgaaga 60

ctgtatccac 70

<210> 238
<211> 70
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement
plasmid sequence

<400> 238
gatcctgctg gagcgacaat ttggagcagt cttcacaacg gccgtgactg agcctcgaag 60

actgtatcac 70

<210> 239
<211> 70
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement
plasmid sequence

<400> 239
gatcctgctg gagcgacaat ttaggagcag tcttccaacg gccgtgactg agcctcagaa 60

gactgtatcc 70

<210> 240
<211> 70
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement
plasmid sequence

<400> 240
gatcctgctg gagcgacaat ttacggagca gtcttcaacg gccgtgactg agcctcacga 60

agactgtatc 70

<210> 241
<211> 70
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: replacement

plasmid sequence

<400> 241
gatcctgctg gagcgacaat ttacaggagc agtcttcacg gccgtgactg agcctcacgg 60

aagactgtat 70

<210> 242
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 242
gatactcaac tctgtctcct tcc 23

<210> 243
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 243
gatactcaac tctgtctcct tcctcttc 28

<210> 244
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 244
gatactcaac tctgtctcct tcctcttcct ac 32

<210> 245
<211> 24

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 245
ggagccccac agctgcacag ggca 24

<210> 246
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 246
ggagccccac agctgcacag ggcaggtct 29

<210> 247
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 247
ggagccccac agctgcacag ggcaggtctt g 31

<210> 248
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 248
gatacgtgca gctgtgggtt gattc 25

<210> 249
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 249
gatacgtgca gctgtgggtt gattccac 28

<210> 250
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 250
gatacgtgca gctgtgggtt gattccacac 30

<210> 251
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 251
ggagccacaa cctccgtcat gtg 23

<210> 252
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 252
ggagccacaa cctccgtcat gtgctgtgac 30

<210> 253
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 253
ggagccacaa cctccgtcat gtgctg 26

<210> 254
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 254
gatacgacgg aggttgtgag gcg 23

<210> 255
<211> 25
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 255
gatacgacgg aggttgtgag gcgct 25

<210> 256
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 256

gatacgacgg aggttgtgag gcgctgcccc cac 33

<210> 257
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 257
ggagcggcaa ccagccctgt cgt 23

<210> 258
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 258
ggagcggcaa ccagccctgt cgtctct 27

<210> 259
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: PCR primer

<400> 259
ggagcggcaa ccagccctgt cgtctctcca 30

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 September 2000 (28.09.2000)

PCT

(10) International Publication Number
WO 00/56923 A3

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/GB00/01128

(22) International Filing Date: 24 March 2000 (24.03.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
9906833.0 24 March 1999 (24.03.1999) GB
9927520.8 23 November 1999 (23.11.1999) GB

(71) Applicant *(for all designated States except US)*: **CLATTERBRIDGE CANCER RESEARCH TRUST** [GB/GB]; J. K. Douglas Laboratories, Clatterbridge Hospital, Bebington, Wirral, Cheshire CH63 4JY (GB).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **SIBSON, Ross** [GB/GB]; One Castlehill Farm Barn, Castlehill, Kingswood, Frodsham, Cheshire WA6 6JS (GB).

(74) Agent: **MCNEIGHT & LAWRENCE**; Regent House, Heaton Lane, Stockport, Cheshire SK4 1BS (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

(88) Date of publication of the international search report:
3 January 2002

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 00/56923 A3

(54) Title: GENETIC ANALYSIS

(57) Abstract: The present invention relates to genetic analysis of nucleic acids, particularly the analysis of the structure and/or sequence of polynucleotides. The invention also relates to the field of oligonucleotide probes, particularly probes in the form of libraries of oligonucleotide fragments. The invention further concerns the construction of oligonucleotide libraries and the methods of their use in the elucidation of structural or sequence information of sample sequences.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 00/01128

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12Q C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, WPI Data, PAJ, MEDLINE, BIOSIS

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 97 29212 A (GINGERAS THOMAS A ;CHEE MARK S (US); STRYER LUBERT (US); AFFYMETRI) 14 August 1997 (1997-08-14) page 4, line 23 -page 5; examples page 23 -page 24 page 29-33 ---	1-10, 18-24,29
X	WO 98 41657 A (CHEE MARK ;AFFYMETRIX INC (US)) 24 September 1998 (1998-09-24) the whole document ---	1-10, 18-24,29
X	WO 95 11995 A (AFFYMAX TECH NV ;FODOR STEPHEN P A (US); GINGERAS THOMAS R (US); L) 4 May 1995 (1995-05-04) page 1-5 page 8 page 26-27 --- -/--	1-10, 18-24,29

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 September 2001	27/09/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Reuter, U

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 00/01128

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	US 5 728 524 A (SIBSON DAVID ROSS) 17 March 1998 (1998-03-17) column 21-24	1-29
A	US 5 846 719 A (BRENNER SYDNEY ET AL) 8 December 1998 (1998-12-08) the whole document	1-29
A	WO 93 17126 A (NEW YORK HEALTH RES INST) 2 September 1993 (1993-09-02) page 2-4 page 25-30; figure 6; examples 2-4	1-29

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

Ir ...national Application No
PCT/GB 00/01128

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 9729212	A	14-08-1997	AU	2189397 A	28-08-1997
			EP	0937159 A1	25-08-1999
			JP	2000504575 T	18-04-2000
			WO	9729212 A1	14-08-1997
			US	6228575 B1	08-05-2001
WO 9841657	A	24-09-1998	EP	0972078 A1	19-01-2000
			WO	9841657 A1	24-09-1998
			WO	9939004 A1	05-08-1999
WO 9511995	A	04-05-1995	AU	8126694 A	22-05-1995
			EP	0730663 A1	11-09-1996
			JP	9507121 T	22-07-1997
			WO	9511995 A1	04-05-1995
			US	6156501 A	05-12-2000
			US	6045996 A	04-04-2000
			US	5861242 A	19-01-1999
			US	5837832 A	17-11-1998
US 5728524	A	17-03-1998	AT	159986 T	15-11-1997
			AU	686563 B2	12-02-1998
			AU	4575893 A	31-01-1994
			CA	2139944 A1	20-01-1994
			DE	69315074 D1	11-12-1997
			DE	69315074 T2	05-03-1998
			EP	0650528 A1	03-05-1995
			WO	9401582 A1	20-01-1994
			JP	7508883 T	05-10-1995
US 5846719	A	08-12-1998	US	5604097 A	18-02-1997
			AU	733782 B2	24-05-2001
			AU	3374097 A	05-01-1998
			CN	1230226 A	29-09-1999
			CZ	9803979 A3	14-07-1999
			EP	0923650 A1	23-06-1999
			HU	0003944 A2	28-03-2001
			JP	2000515006 T	14-11-2000
			NO	985698 A	08-02-1999
			PL	331513 A1	19-07-1999
			US	6138077 A	24-10-2000
			US	6172218 B1	09-01-2001
			WO	9746704 A1	11-12-1997
			US	6235475 B1	22-05-2001
			US	6172214 B1	09-01-2001
			US	6150516 A	21-11-2000
			US	6013445 A	11-01-2000
			AU	712929 B2	18-11-1999
			AU	4277896 A	06-05-1996
			AU	5266399 A	09-12-1999
			CA	2202167 A1	25-04-1996
			CZ	9700866 A3	17-09-1997
			EP	0793718 A1	10-09-1997
			FI	971473 A	04-06-1997
			HU	77916 A2	28-10-1998
			JP	10507357 T	21-07-1998
			NO	971644 A	02-06-1997
			US	6280935 B1	28-08-2001
			WO	9612014 A1	25-04-1996

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/GB 00/01128

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5846719 A		US 5635400 A	03-06-1997
		US 5654413 A	05-08-1997
		AU 3946195 A	06-05-1996
		DE 69513997 D1	20-01-2000
		DE 69513997 T2	27-07-2000
		EP 0786014 A1	30-07-1997
		EP 0952216 A2	27-10-1999
		WO 9612039 A1	25-04-1996
		US 6140489 A	31-10-2000
		US 5695934 A	09-12-1997
		US 5863722 A	26-01-1999
WO 9317126 A	02-09-1993	AU 3728093 A	13-09-1993
		CA 2130562 A1	02-09-1993
		EP 0675966 A1	11-10-1995
		WO 9317126 A1	02-09-1993
		US 6103463 A	15-08-2000